OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 17:34:09 ; Search time 4339 Seconds
                 (without alignments)
                 11028.050 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaaactaggcccta.........gtgtcattgctttatgact 1104

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_nam:*
                 37: em_htg_vrt:*
                 38: em_sy:*
                 39: em_htgo_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1104	100.0	1104	6	AX702273	AX702273 Sequence
2	779	70.6	4319	6	AX647101	AX647101 Sequence
3	779	70.6	98861	9	AL136961	Human DNA
4	551	49.9	151174	2	AC016189	Homo sapi
5	397.2	36.0	536	6	AX230148	Sequence
6	378.4	34.3	556	6	AX244735	Sequence
7	353.4	32.0	151174	2	AC016189	Homo sapi
8	261.2	23.7	162543	2	BX470183	Danio rer
9	221.6	20.1	190588	2	AC102718	Mus muscu
10	221.6	20.1	204438	10	AC110249	Mus muscu
11	220	19.9	1493	6	BC052868	Mus muscu
12	208.6	18.9	239344	2	AC133310	Rattus no
13	208.6	18.9	244625	2	AC094739	Rattus no
14	184.8	16.7	1086	6	AR308611	Sequence
15	184.8	16.7	1086	6	AX704977	Sequence
16	184.8	16.7	1643	6	AR003350	Sequence
17	184.8	16.7	1643	6	AR269382	Sequence
18	184.8	16.7	1643	6	AX270716	Sequence
19	184.8	16.7	1643	6	AX774793	Sequence
20	184.8	16.7	1643	6	HUMGPCRB	Sequence
21	183.2	16.6	1086	6	AR308705	Sequence
22	183.2	16.6	1638	6	AX548826	Sequence
23	183.2	16.6	1710	9	BC020752	Homo sapi
24	183.2	16.6	149478	2	AL160155	Human DNA
25	165	14.9	187201	4	BX004885	Zebrafish
26	119.6	10.8	1238	4	OCU33334	Oryctolagus
27	119.4	10.8	1020	6	AX458308	Sequence
28	119.4	10.8	1020	6	BD187417	Novel pol
29	119.4	10.8	1076	6	AX375230	Sequence
30	119.4	10.8	2025	6	AR270919	Sequence
31	118.6	10.7	2025	6	AX548931	Sequence
32	118.6	10.7	2025	9	HS007225	Homo sapien
33	183.2	16.6	1382	4	AB051422	Sus scrof
34	117.8	10.7	184377	10	AC121579	Mus muscu
35	117.8	10.7	195290	2	AC135635	Mus muscu
36	117	10.6	1134	9	AY136753	Homo sapi
37	117	10.6	2257	9	BC028135	Homo sapi
38	117	10.6	2681	9	BC012104	Homo sapi
39	117	10.6	123185	9	AP000587	Homo sapi
40	116.4	10.5	196988	9	AP002761	Homo sapi
41	116.4	10.5	250610	2	AC120742	Rattus no
42	113.4	10.3	276127	2	AC116265	Rattus no
43	113.4	10.3	2025	6	AR308577	Sequence
44	112	10.1	1020	6	AR308693	Sequence
45	112	10.1	1741	9	HS294154	H.sapiens m

ALIGNMENTS

RESULT 1
AX702273

LOCUS   AX702273   1104 bp   DNA   linear   PAT 03-APR-2003
DEFINITION   Sequence 33 from Patent WO02064793.
ACCESSION   AX702273
VERSION   AX702273.1 GI:29537547
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Casman,S.J., Edinger,S.R., Ellerman,K., Smithson,G., Kekuda,R. and
                Muraidhara,P.
TITLE   Novel gpcr-like proteins and nucleic acids encoding same

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Patent: WO 02064793-A 33 22-AUG-2002;
             Curagen Corporation (US)
FEATURES   Location/Qualifiers
     source   1..1104
                 /organism="Homo sapiens"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:9606"
     CDS   1..1083
                 /codon_start=1
                 /note="unnamed protein product"
                 /protein_id="CAD87717.1"
                 /db_xref="GI:29537548"
                 /db_xref="REMTREMBL:CAD87717"
                 /translation="MIKLGPYFTPFKIKIKDLNVESQTVKIFLRSLFYTALLVFS
                 ALGNILALCLTCQKSRKINCTGIVLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGL
                 CRLTAFVLYDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL
                 QTMPLLIMPMTKPLVGKLLACMEYSSMESVLGLPLMVLVAPAIGFCGPVGIILSCYMKI
                 TWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVCFSPYHLNIKQPWARGMLHLPS
                 CABRRAFLLSLQATVALMNNNCGIITPIIYFFASTHYRKWILLGILKLKGSSSSSSSSSS
                 TPGKASSETPSITQARGSMFLAEHVV"
ORIGIN

Query Match   100.0%; Score 1104; DB 6; Length 1104;
Best Local Similarity 100.0%; Pred. No. 2.3e-215;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGATAAAACTAGGCCCTTTACTTCACACCACCACCACCACACAAAAAAAATAAAGATTAAAGAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGATAAAACTAGGCCCTTTACTTCACACCACCACCACCACACAAAAAAAATAAAGATTAAAGAT 60

Qy   61 TTAAATGTGGAAAGCCAAACTGTGAAAAATTTTTTTTGGCTCTTCTGTTCTTGTTCTTGTTCTTCTACACAGCCCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTAAATGTGGAAAGCCAAACTGTGAAAAATTTTTTTTGGCTCTTCTGTTCTTGTTCTTGTTCTTCTACACAGCCCTC 120

Qy   121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGAGCAGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGAGCAGG 180

Qy   181 AAGATCAACTGCACAGGCATCTACCTGGGTGCACCTGGGTGTCCTGGGTGTGTTCTGACCTGTGTTCACC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGATCAACTGCACAGGCATCTACCTGGGTGCACCTGGGTGTCCTGGGTGTGTTCTGACCTGTGTTCACC 240

Qy   241 GTGCGCCTTACCGGGAAAGGGGTGGGTGTGTTATGTGCTGGGCTCCTTCGGGCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTGCGCCTTACCGGGAAAGGGGTGGGTGTGTTATGTGCTGGGCTCCTTCGGGCAAG 300

Qy   301 GGGCTCTGCAGGCTGACACGGCGTTTGTGCTTCTACACGGACCTTACCTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGGCTCTGCAGGCTGACACGGCGTTTGTGCTTCTACACGGACCTTACCTC 360

Qy   361 ATGGCCCTGTGTGAGCGGTGACCATTACCCAGCCGTTGCCCACTGGGGCCCCGGGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ATGGCCCTGTGTGAGCGGTGACCATTACCCAGCCGTTGCCCACTGGGGCCCCGGGC 420

Qy   421 CTCCCGCACGGCGCCGGCGCAGGCTGTGTCTGCGGTGGCCATCTTGGACCTTGGTGCTCGTG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTCCCGCACGGCGCCGGCGCAGGCTGTGTCTGCGGTGGCCATCTTGGACCTTGGTGCTCGTG 480

Qy   481 CAGACGAATGCCCTTGCTCTTGATGCCCATGACCAAGCGCCGCAAGCTGGCCTGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAGACGAATGCCCTTGCTCTTGATGCCCATGACCAAGCGCCGCAAGCTGGCCTGC 540

Qy   541 ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGTGCTCCTTGGTGCTGTGGCCTTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGTGCTCCTTGGTGCTGTGGCCTTT 600

Qy   601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCTTGCTATATGAAAGATCAACCTGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCTTGCTATATGAAAGATCAACCTGG 660

Qy   661 AAGCTGTGCAGCACAGCTCGGGAGGACACCCAGTGACCAGCAGGAGAAAGATCTTCTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AAGCTGTGCAGCACAGCTCGGGAGGACACCCAGTGACCAGCAGGAGAAAGATCTTCTT 720

Qy   721 GGCTGCCTGCTTACCGCTGCTGAGTGCTGGTGGTCCGTGGTGTTGGTGTTCTGCTTCAGCCCTTACCAC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGCTGCCTGCTTACCGCTGCTGAGTGCTGGTGGTCCGTGGTGTTGGTGTTCTGCTTCAGCCCTTACCAC 780

Qy   781 CTCAACATCAAGCAGTCAATGGGCAGAGAGGAGATGCTCAACCTGCCATCCTGTTGCCCGAGCGCGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CTCAACATCAAGCAGTCAATGGGCAGAGAGGAGATGCTCAACCTGCCATCCTGTTGCCCGAGCGCGG 840

Qy   841 AGGGGCTTTCTTACTGTCCTTCAGGCCACCGTGGCGCCTCATGAACATGAACTGTGGGCATT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AGGGGCTTTCTTACTGTCCTTCAGGCCACCGTGGCGCCTCATGAACATGAACTGTGGGCATT 900

Qy   901 ACCCCAATCAATTACTTACTTCTTTGCATCCACCCCATTACACAGGAGAAATGGCTTCCTGGCCATTTTA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 ACCCCAATCAATTACTTACTTCTTTGCATCCACCCCATTACACAGGAGAAATGGCTTCCTGGCCATTTTA 960

Qy   961 AAGCTCAAAAGGGTCTTCCTCCTCCTCCTCCTCCACCCACCCCAGAGAAAAGCTTCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGCTCAAAAGGGTCTTCCTCCTCCTCCTCCTCCACCCACCCCAGAGAAAAGCTTCT 1020

Qy   1021 TCAGAAAACACCAAAGTATCACCCAGCCCAGAGAGGTTCTTATGTTCTTAGGCGGCACCGTGGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TCAGAAAACACCAAAGTATCACCCAGCCCAGAGAGGTTCTTATGTTCTTAGGCGGCACCGTGGTC 1080

Qy   1081 TAACGTGTCATTTGCTTTATGACT 1104
          ||||||||||||||||||||||||
Db   1081 TAACGTGTCATTTGCTTTATGACT 1104

RESULT 2
AX647101
LOCUS   AX647101   4319 bp   DNA   linear   PAT   04-MAR-2003
DEFINITION Sequence 1293 from Patent EP1270724.
ACCESSION AX647101
VERSION   AX647101.1 GI:28800143
KEYWORDS
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
  AUTHORS Suwa,M., Asai,K., Akiyama,Y. and Aburatani,H.
  TITLE   Guanosine triphosphate-binding protein coupled receptors
  JOURNAL Patent: EP 1270724-A 1293 02-JAN-2003;
           National Institute of Advanced Industrial Science and Technology
           (JP) ; Center for Advanced Science and Technology Incubation, Ltd.
           (JP)
FEATURES   Location/Qualifiers
     source   1..4319
                 /organism="Homo sapiens"
                 /mol_type="unassigned DNA"
                 /db_xref="taxon:9606"
     CDS   join(201..308,3128..3531,3660..4119)
                 /codon_start=1
                 /note="unnamed protein product"
                 /protein_id="CAD69676.1"
                 /db_xref="GI:28800144"
                 /db_xref="REMTREMBL:CAD69676"
                 /translation="MFVDSVGQDIEEGTVGMIFLCSLTLEVSARKDQQIGLAFRQGAL
                 QADGVCALRHRHLRGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQ
                 TMPLLIMPMTKPLVGKLLACMEYSSMESVLGLPLMVLVAPAIGFCGPVGIILSCYMKIT
                 WKLCSTAGRTQTAREDPVISGKGRHRRDSPGGPSDQQERTPLARLPAYAADAGRGGL
                 LQPLPEPQAVHGERDAPPAILCRAEGFIVPSGHRGPHEHELWHYPIIYFFASTHYR
                 KWLLGTLKLKGSSSSSSTPGKASSETPSITQARGSMFLAEHVV"
ORIGIN

Query Match   70.8%; Score 779; DB 6; Length 4319;
Best Local Similarity 83.6%; Pred. No. 6.9e-149;
Matches 1006; Conservative 0; Mismatches 0; Indels 197; Gaps 3;

Qy   97 CGGCTGCTTCTGTTCTTACACAGACCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Qy   1042 CAGGCCAGAGGTTCTATGTTCTTAGCCGAGACACGTGGTGTCTAACGTGGTTCATTTGTTT 1101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4078 CAGGCCAGAGGTTCTATGTTCTTAGCCGAGACACGTGGTGTCTAACGTGGTTCATTTGTTT 4137

Qy   1102 ACT 1104
          |||
Db   4138 ACT 4140
```

RESULT 3

<pre>
LOCUS   AL136961/c   98861 bp   DNA   linear   PRI 11-APR-2001
DEFINITION Human DNA sequence from clone RP11-178C10 on chromosome 13,
            complete sequence.
ACCESSION   AL136961
VERSION   AL136961.19 GI:13620283
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 98861)
  AUTHORS   Blakey,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On Apr 12, 2001 this sequence version replaced gi:13276997.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-178C10 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-178C10 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-178C10 is at 98861 in this
            sequence. The true right end of clone RP11-461N23 is at 100 in this
            sequence.
            The true right end of clone RP11-178C10 is at 100 in this
            sequence.
FEATURES   Location/Qualifiers
     source   1..98861
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-178C10"
                     /clone_lib="RPCI-11.1"
     repeat_region   161..463
                     /note="L1MB8 repeat: matches 5836..6160 of consensus"
     repeat_region   656..723
                     /note="L2 repeat: matches 2592..2659 of consensus"
</pre>

```
repeat_region    2385..2420
                 /note="MIR repeat: matches 111 ..146 of consensus"
repeat_region    4555..4646
                 /note="L2 repeat: matches 1031 ..1120 of consensus"
repeat_region    4989..5306
                 /note="MER7A repeat: matches 2 ..345 of consensus"
repeat_region    5859..5987
                 /note="L2 repeat: matches 2574 ..2708 of consensus"
repeat_region    6215..6548
                 /note="Charlie1 repeat: matches 2292 ..2629 of consensus"
repeat_region    6557..6635
                 /note="L2 repeat: matches 2128 ..2206 of consensus"
repeat_region    6730..7233
                 /note="L2 repeat: matches 1330 ..1889 of consensus"
repeat_region    8301..8499
                 /note="MIR repeat: matches 17 ..219 of consensus"
repeat_region    8893..8994
                 /note="L2 repeat: matches 1551 ..1656 of consensus"
repeat_region    9079..9126
                 /note="24 copies 2 mer at 75% conserved"
repeat_region    9554..9754
                 /note="MER20 repeat: matches 33 ..218 of consensus"
repeat_region    9816..9979
                 /note="FAM repeat: matches 2 ..165 of consensus"
repeat_region    10402..10703
                 /note="L1M1 repeat: matches 5264 ..5577 of consensus"
repeat_region    12181..12626
                 /note="L1 repeat: matches 3474 ..3949 of consensus"
repeat_region    12654..13012
                 /note="L1ME3 repeat: matches 5571 ..5927 of consensus"
repeat_region    13491..14076
                 /note="L1 repeat: matches 2226 ..2881 of consensus"
repeat_region    16310..17185
                 /note="L1MC4 repeat: matches 6237 ..7151 of consensus"
repeat_region    17207..17622
                 /note="MER47B repeat: matches 2 ..413 of consensus"
repeat_region    17628..18380
                 /note="L1MC4 repeat: matches 7197 ..7976 of consensus"
repeat_region    20797..21100
                 /note="AluY repeat: matches 1 ..297 of consensus"
repeat_region    22746..22814
                 /note="L2 repeat: matches 2574 ..2643 of consensus"
repeat_region    23758..23981
                 /note="AluJb repeat: matches 85 ..310 of consensus"
repeat_region    25108..25512
                 /note="L2 repeat: matches 2570 ..2688 of consensus"
repeat_region    25740..25795
                 /note="MIR repeat: matches 196 ..251 of consensus"
repeat_region    25982..26012
                 /note="Tigger4 (Zombi) repeat: matches 2701 ..2731 of
                 consensus"
repeat_region    26005..26208
                 /note="Tigger4 (Zombi) repeat: matches 1772 ..1970 of
                 consensus"
repeat_region    26209..26513
                 /note="Tigger4 (Zombi) repeat: matches 1 ..1772 of consensus"
repeat_region    26514..28201
                 /note="AluSx repeat: matches 1 ..305 of consensus"
repeat_region    32616..32843
                 /note="6 copies 38 mer 86% conserved"
repeat_region    35333..35362
                 /note="15 copies 2 mer ac 100% conserved"
repeat_region    36220..36303
                 /note="2 copies 42 mer 97% conserved"
repeat_region    36967..37274
                 /note="AluSg repeat: matches 1 ..313 of consensus"
repeat_region    39481..39767
                 /note="AluSx repeat: matches 1 ..297 of consensus"
repeat_region    41084..41111
                 /note="14 copies 2 mer aa 89% conserved"
repeat_region    41119..41174
                 /note="THE1C repeat: matches 1 ..58 of consensus"

repeat_region    41175..41482
                 /note="AluSp repeat: matches 1 ..309 of consensus"
repeat_region    41483..41785
                 /note="THE1C repeat: matches 58 ..344 of consensus"
repeat_region    41791..41842
                 /note="L2 repeat: matches 1215 ..1266 of consensus"
repeat_region    42051..42573
                 /note="MER74A repeat: matches 2 ..555 of consensus"
repeat_region    42857..43014
                 /note="L2 repeat: matches 2553 ..2733 of consensus"
repeat_region    43250..43279
                 /note="15 copies 2 mer ca 100% conserved"
repeat_region    44262..44708
                 /note="MLT1C repeat: matches 1 ..466 of consensus"
repeat_region    44860..45175
                 /note="MLT2FB repeat: matches 1 ..328 of consensus"
repeat_region    45204..45235
                 /note="16 copies 2 mer tg 87% conserved"
repeat_region    45389..45559
                 /note="MER5B repeat: matches 4 ..175 of consensus"
repeat_region    45639..45788
                 /note="MER5A repeat: matches 9 ..184 of consensus"
repeat_region    45963..46162
                 /note="MIR repeat: matches 25 ..250 of consensus"
repeat_region    46281..46837
                 /note="MER67D repeat: matches 1 ..510 of consensus"
repeat_region    47886..51275
                 /note="L1PA6 repeat: matches 2744 ..6141 of consensus"
repeat_region    51364..51523
                 /note="MER5A repeat: matches 1 ..189 of consensus"
repeat_region    51736..51791
                 /note="MIR repeat: matches 118 ..178 of consensus"
repeat_region    51848..52167
                 /note="MSTB repeat: matches 86 ..422 of consensus"
repeat_region    53138..53915
                 /note="MER21B repeat: matches 41 ..789 of consensus"
repeat_region    54321..54629
                 /note="AluSg repeat: matches 1 ..310 of consensus"
repeat_region    55247..55308
                 /note="MER69A repeat: matches 24 ..87 of consensus"
repeat_region    55309..55622
                 /note="AluSg repeat: matches 1 ..313 of consensus"
repeat_region    55623..55715
                 /note="MER69A repeat: matches 87 ..176 of consensus"
repeat_region    57171..57271
                 /note="MLT1F repeat: matches 437 ..541 of consensus"
repeat_region    57462..57560
                 /note="MLT1G repeat: matches 169 ..264 of consensus"
repeat_region    57561..58049
                 /note="MER74A repeat: matches 1 ..519 of consensus"
repeat_region    58063..58287
                 /note="AluSg/x repeat: matches 163 ..300 of consensus"
repeat_region    58336..58475
                 /note="MLT1G repeat: matches 76 ..205 of consensus"
repeat_region    58496..58809
                 /note="AluSx repeat: matches 1 ..311 of consensus"
repeat_region    58818..59182
                 /note="L1ME3A repeat: matches 5789 ..6157 of consensus"
repeat_region    59198..59336
                 /note="MER74A repeat: matches 1 ..139 of consensus"
repeat_region    59337..59435
                 /note="MLT1G repeat: matches 169 ..264 of consensus"
repeat_region    59626..59726
                 /note="MLT1F repeat: matches 437 ..541 of consensus"
repeat_region    59804..60249
                 /note="L1ME3A repeat: matches 5301 ..5782 of consensus"


Query Match            70.6%;   Score 779;  DB 9;  Length 98861;
Best Local Similarity  83.6%;   Pred. No. 7.6e-149;
Matches 1006;  Conservative   0;  Mismatches   0;  Indels 197;  Gaps   3;

Qy    97 CGCTCTCGTTCTACACAGCCCTCCTGGTTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   1042  CAGGCCAGAGGTTCTATGTTCTTAGCGGAGCAGCTGGTCTAACGTGTCATTGCTTTATG 1101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55784  CAGGCCAGAGGTTCTATGTTCTTAGCGGAGCAGCTGGTCTAACGTGTCATTGCTTTATG 55725

Qy   1102 ACT 1104
           |||
Db  55724 ACT 55722
```

RESULT 4
AC016189

```
LOCUS       AC016189              151174 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016189
VERSION     AC016189.3  GI:9129134
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151174)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-23M9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151174)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,K., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,J., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6939915.
            All repeats were identified using RepeatMasker.
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L4399
            Center clone name: 23_M_9
            ---------------
            * NOTE: This record contains 152 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1     851: contig of 851 bp in length
            *      852     951: gap of 100 bp
            *      952    1804: contig of 853 bp in length
            *     1805    1904: gap of 100 bp
            *     1905    2719: contig of 815 bp in length
            *     2720    2819: gap of 100 bp
```

```
Db  56922  CGCTCCTCGTCTACACAGCCCTCCTGGTCTTCAGTCGCCCTGGGAAACATCCTTGCCCTT 56863

Qy   157   TGCCTTACCTGTCAAAAGAGCAGGAGGAGATCAACTGCACAGGCATCTACCTGGTCACCTG 216
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56862  TGCCTTACCTGTCAAAAGAGCAGGAGGAGATCAACTGCACAGGCATCTACCTGGTCACCTG 56803

Qy   217   GCTGTGTCTGACCTGCTGTTCACCGTGGCCTGGCCTTACCGGGAAGGGTGTGGTTATGTGCTG 276
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56802  GCTGTGTCTGACCTGCTGTTCACCGTGGCCTGGCCTTACCGGGAAGGGTGTGGTTATGTGCTG 56743

Qy   277   GGTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGCGGTTTGTGCTCTACACC 336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56742  GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGCGGTTTGTGCTCTACACC 56683

Qy   337   GACACCTACGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 396
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  56682  GACACCTAC-GGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 56624

Qy   397   GTCTGTGCCCACTGGGGGCCGCCGCCTCCGCACGGCTGCCGCCCAGGCTGCTCTCGCTG 456
           ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  56623  GTCTGTGCCCACTGGGGGCCCGCCGCCTCCGCACGGCTGCCGCCCAGGCTGCTCTCGCTG 56564

Qy   457   GCCATCTGGACCTGGTGCTGCTGCAGACGATGCCCTGCTCTTGATGCCCATGACCAAG 516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56563  GCCATCTGGACCTGGTGCTGCTGCAGACGATGCCCTGCTCTTGATGCCCATGACCAAG 56504

Qy   517   CCGCTGGTGGGCAAGCTGGCCTGCCATGGAGTACAGCAGCATGAGTCAGTCCTCGGGCTG 576
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56503  CCGCTGGTGGGCAAGCTGGCCTGCCATGGAGTACAGCAGCATGAGTCAGTCCTCGGGCTG 56444

Qy   577   CCCCTCATGGTCTCCTGTGCCTTTGCCATTGGCTTCTGTGGGGCCAGTGGGGATCATCCTG 636
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56443  CCCCTCATGGTCTCCTGTGCCTTTGCCATTGGCTTCTGTGGGGCCAGTGGGGATCATCCTG 56384

Qy   637   TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT------------ 678
           |||||||||||||||||||||||||||||||||||||||||
Db  56383  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCTGGAGAGACCTCGGGAGACCA 56324

Qy   679   ------------ 678

Db  56323  GCGGGAAAGGACGACACCACCGGCGGGGCGGCCCAGGAGACCCAGTGACCAACAGGAACGA 56264

Qy   679   ------------ 678

Db  56263  CGCCACCCGGCGGGGGGGGCGGCCCAGGAGACCCAGTGACCAGCGGGAAAGGACACCACCCGGCA 56204

Qy   679   ------------CGG 681

Db  56203  GGACAGCCCGAGAGGACCCCAGTGACCAGGAAAGAGCGCCACCCGGCGGGGACAGCCCGG 56144

Qy   682   GAGGACCCAGTGACCAGCAGGAAAGGACGCGCCACTGGCGAGGTCTGCCTGCTTACGCTGCTG 741
           ||||||||||||||||||||||||||||||||||||||||||
Db  56143  GAGGACCCAGTGACCAGCAGGAAAGGACGCGCCACTGGCGAGGCTGCCTGCTTACGCTGCTG 56084

Qy   742   ATGCTGGTGGCCGTGGCCGTGGTGGTCTGCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATG 801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56083  ATGCTGGTGGCCGTGGCCGTGGTGGTCTGCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATG 56024

Qy   802   GCGGAGGGGATGCTCCACCTGGCCATCCTGTGCCGGAGCGGAGGGGCTTCTTCTACTGTCCCTT 861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56023  GCGGAGGGGATGCTCCACCTGGCCATCCTGTGCCGGAGCGGAGGGGCTTCTTCTACTGTCCCTT 55964

Qy   862   CAGGCCACCGTGGCCCTGGCCCTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTCTTT 921
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  55963  CAGGCCACCGTGGCCCTGGCCCTCATGAACATGAACTGTGGCATTACCCC-ATCATTACTTCTTT 55905

Qy   922   GCATCCACCCATTACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCC 981
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55904  GCATCCACCCATTACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCC 55845

Qy   982   TCCTCCTCCTCCTCCTCCTCCTCCACCCCACCCCACCCCAGAAACACCAAGTATCACC 1041
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55844  TCCTCCTCCTCCTCCTCCTCCTCCACCCCACCCCACCCCAGAAACACCAAGTATCACC 55785
```

```
*   2820   3668: contig of 849 bp in length
*   3669   3768: gap of 100 bp
*   3769   4657: contig of 889 bp in length
*   4658   4757: gap of 100 bp
*   4758   5600: contig of 843 bp in length
*   5601   5700: gap of 100 bp
*   5701   6532: contig of 832 bp in length
*   6533   6632: gap of 100 bp
*   6633   7481: contig of 849 bp in length
*   7482   7581: gap of 100 bp
*   7582   8447: contig of 866 bp in length
*   8448   8547: gap of 100 bp
*   8548   9383: contig of 836 bp in length
*   9384   9483: gap of 100 bp
*   9484  10342: contig of 859 bp in length
*  10343  10442: gap of 100 bp
*  10443  11323: contig of 881 bp in length
*  11324  11423: gap of 100 bp
*  11424  12254: contig of 831 bp in length
*  12255  12354: gap of 100 bp
*  12355  13288: contig of 934 bp in length
*  13289  13388: gap of 100 bp
*  13389  14235: contig of 847 bp in length
*  14236  14335: gap of 100 bp
*  14336  15172: contig of 837 bp in length
*  15173  15272: gap of 100 bp
*  15273  16130: contig of 858 bp in length
*  16131  16230: gap of 100 bp
*  16231  17073: contig of 843 bp in length
*  17074  17173: gap of 100 bp
*  17174  18003: contig of 830 bp in length
*  18004  18103: gap of 100 bp
*  18104  18939: contig of 836 bp in length
*  18940  19039: gap of 100 bp
*  19040  19885: contig of 846 bp in length
*  19886  19985: gap of 100 bp
*  19986  20834: contig of 849 bp in length
*  20835  20934: gap of 100 bp
*  20935  21761: contig of 827 bp in length
*  21762  21861: gap of 100 bp
*  21862  22743: contig of 882 bp in length
*  22744  22843: gap of 100 bp
*  22844  23708: contig of 865 bp in length
*  23709  23808: gap of 100 bp
*  23809  24676: contig of 868 bp in length
*  24677  24776: gap of 100 bp
*  24777  25602: contig of 826 bp in length
*  25603  25702: gap of 100 bp
*  25703  26560: contig of 858 bp in length
*  26561  26660: gap of 100 bp
*  26661  27504: contig of 844 bp in length
*  27505  27604: gap of 100 bp
*  27605  28449: contig of 845 bp in length
*  28450  28549: gap of 100 bp
*  28550  29398: contig of 849 bp in length
*  29399  29498: gap of 100 bp
*  29499  30357: contig of 859 bp in length
*  30358  30457: gap of 100 bp
*  30458  31323: contig of 866 bp in length
*  31324  31423: gap of 100 bp
*  31424  32300: contig of 877 bp in length
*  32301  32400: gap of 100 bp
*  32401  33254: contig of 854 bp in length
*  33255  33354: gap of 100 bp
*  33355  34218: contig of 864 bp in length
*  34219  34318: gap of 100 bp
*  34319  35169: contig of 851 bp in length
*  35170  35269: gap of 100 bp
*  35270  36125: contig of 856 bp in length
*  36126  36225: gap of 100 bp
*  36226  37061: contig of 836 bp in length
*  37062  37161: gap of 100 bp
*  37162  38003: contig of 842 bp in length

*  38004  38103: gap of 100 bp
*  38104  38945: contig of 842 bp in length
*  38946  39045: gap of 100 bp
*  39046  39882: contig of 837 bp in length
*  39883  39982: gap of 100 bp
*  39983  40833: contig of 851 bp in length
*  40834  40933: gap of 100 bp
*  40934  41779: contig of 846 bp in length
*  41780  41879: gap of 100 bp
*  41880  42722: contig of 843 bp in length
*  42723  42822: gap of 100 bp
*  42823  43678: contig of 856 bp in length
*  43679  43778: gap of 100 bp
*  43779  44634: contig of 856 bp in length
*  44635  44734: gap of 100 bp
*  44735  45587: contig of 853 bp in length
*  45588  45687: gap of 100 bp
*  45688  46542: contig of 855 bp in length
*  46543  46642: gap of 100 bp
*  46643  47488: contig of 846 bp in length
*  47489  47588: gap of 100 bp
*  47589  48448: contig of 860 bp in length
*  48449  48548: gap of 100 bp
*  48549  49450: contig of 902 bp in length
*  49451  49550: gap of 100 bp
*  49551  50383: contig of 833 bp in length
*  50384  50483: gap of 100 bp
*  50484  51353: contig of 870 bp in length
*  51354  51453: gap of 100 bp
*  51454  52306: contig of 853 bp in length
*  52307  52406: gap of 100 bp
*  52407  53252: contig of 846 bp in length
*  53253  53352: gap of 100 bp
*  53353  54224: contig of 872 bp in length
*  54225  54324: gap of 100 bp
*  54325  55162: contig of 838 bp in length
*  55163  55262: gap of 100 bp
*  55263  56086: contig of 824 bp in length
*  56087  56186: gap of 100 bp
*  56187  57026: contig of 840 bp in length
*  57027  57126: gap of 100 bp
*  57127  58025: contig of 899 bp in length
*  58026  58125: gap of 100 bp
*  58126  59004: contig of 879 bp in length
*  59005  59104: gap of 100 bp
*  59105  59979: contig of 875 bp in length
*  59980  60079: gap of 100 bp
*  60080  60920: contig of 841 bp in length
*  60921  61020: gap of 100 bp
*  61021  61885: contig of 865 bp in length
*  61886  61985: gap of 100 bp
*  61986  62792: contig of 807 bp in length
*  62793  62892: gap of 100 bp
*  62893  63744: contig of 852 bp in length
*  63745  63844: gap of 100 bp
*  63845  64693: contig of 849 bp in length
*  64694  64793: gap of 100 bp
*  64794  65645: contig of 852 bp in length
*  65646  65745: gap of 100 bp
*  65746  66624: contig of 879 bp in length
*  66625  66724: gap of 100 bp
*  66725  67550: contig of 826 bp in length
*  67551  67650: gap of 100 bp
*  67651  68502: contig of 852 bp in length
*  68503  68602: gap of 100 bp
*  68603  69435: contig of 833 bp in length
```

```
Query Match      49.9%;   Score 551;   DB 2;   Length 151174;
Best Local Similarity  97.8%;   Pred. No. 3.2e-102;
Matches  590;   Conservative  0;   Mismatches  10;   Indels  3;   Gaps  3;

Qy    126 CTTCAGTGCCCTGGGAAACATCCTTGCCCTTGCCTTACCTGTCA-AAAGAGCAGGAAGA  184
          ||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
```

```
Db    138636 CTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAGAGAGGAGGAGAAGA 138695
                                                                                Db    345 CGGGGGGGTCTCACCTCATGGCCTGTCTGTGGAGCGTGGACCATTACCCAGCTGTGGTCTGTGTGC 404

Qy       185 TCAACTGCACAGGCATCATCCTGTGCACCTGGTGTGTCTGACCTGTGTTCACCGTGG 244
                                                                                Qy     68 C-GGGGGGTCTCACCTCATGGCCTGTCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGC 126

Db    138696 TCAACTGCACAGGCATCATCCTGTGCACCTGGTGTGTCTGACCTGTGTTCACCGTGG 138755
                                                                                Db    405 CCACTGGGGCCCGCCTCCGCACGCGCTGGCCGCGCGGCTGGTCTGCGGTGGCCATCTG 464

Qy       245 CCTTACCCGGGAAAGGTGGTGTGTTATGTCGGGCTCCAGCTGCCTTTCGGCAAGGGGC 304
                                                                                Qy    127 CCACTGGGGCCCGCCTCCGCACGCGCTGGCCGCGCGGCTGGTCTGCGGTGGCCATCTG 186

Db    138756 CCTTACCCGGGAAAGGTGGTGTGTTATGTCGGGCTCCAGCTGCCTTTCGGCAAGGGGC 138815
                                                                                Db    465 GACCTTGGTGCTGCCAGAGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGT 524

Qy       305 TCTGCAGGCTGACGGCGTTTGTGCTCTACACCACGCACCACCTACCTCATGG 364
                                                                                Qy    187 GACCTTGGTGCTGCCAGAGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGT 246

Db    138816 TCTGCAGGCTGACGGCGTTTGTGCTCTACACCACGCACCACCTACCTCATGG 138874
                                                                                Db    525 GGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGCAGTCAGTCCTC-GGGCTGCCCTCA 583

Qy       365 CCTGTGTGAGCGTGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGGCCCGCCTCC 424
                                                                                Qy    247 GGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGCAGTCAGTCCTCGGGGTGCCCCTCA 306

Db    138875 CCTGTGTGAGCGTGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGGCCCGCCTCC 138934
                                                                                Db    584 TGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCTGCT 643

Qy       425 GCACCGGCTGGCCGCGCGGGCCCAGGCTGGTCTGCGGTCGCCATCTGGACCTTGGTGCTGCTGCTGCAGA 484
                                                                                Qy    307 TGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCTGCT 366

Db    138935 GCACCGGCTGGCCGCGCGGGCCCAGGCTGGTCTGCGGTCGCCATCTGGACCTTGGTGCTGCTGCTGCAGA 138994
                                                                                Db    644 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGACACAGCACCCAGTGACCAGCACGA 703

Qy       485 CGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCCTGCATGG 544
                                                                                Qy    367 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTGGGAGAACCCAGTGACCAGCACGA 425

Db    138995 CGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGCCTGCATGG 139054

Qy       545 AGTACAGCAGCATGGCAGTCAGTCCTCGGGGTGCCCCTCATGGTCCTTTGCCA 604
                                                                                Qy    704 AAGGACGGCCACTGGCGAGGCTGCCTG 729

Db    139055 AGTACAGCAGCATGGCAGTCAGTCCTCGGGGCTGCCCTCATGGTCCTTTGCCA 139114
                                                                                Db    426 AAGGACACCACCGGCGGCGGGCAGCCCG 451

Qy       605 TTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCGCTGTATATGAAGATCACCTGGAAGC 664

Db    139115 TTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCGCTGTATATGAAGATCACCTGGAAGC 139174

Qy       665 TGTGCAGCACAGCTGGGAGGACCCCAGTGACCAGCAGGAAAGGACCTGGCGAGCT

Db    139175 TGGGCAGCACAGCT-GGGAGAACCCAGTGACCAGCAGGGGGAAAAGACACCACCGGCGGCGGGGCA 139233

Qy       725 GCC 727
             |||
Db    139234 GCC 139236
```

RESULT 5

```
AX230148
LOCUS       AX230148                  536 bp    DNA      linear   PAT 11-SEP-2001
DEFINITION  Sequence 35 from Patent WO0162797.
ACCESSION   AX230148
VERSION     AX230148.1   GI:15592169
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Vogeli,G., Wood,L.S., Parodi,L.A. and Lind,P.
  TITLE     Novel g protein-coupled receptors
  JOURNAL   Patent: WO 0162797-A 35 30-AUG-2001;
            PHARMACIA & UPJOHN COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..536
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             36.0%;  Score 397.2;  DB 6;  Length 536;
Best Local Similarity   97.5%;  Pred. No. 7.7e-71;
Matches  435;  Conservative  0;  Mismatches  8;  Indels  3;  Gaps  3;

Qy       285 CTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACACCTA 344
             |||||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db         8 CTGGCCTTTCGGCAAGGGGCTTGCAGGCTGACGGCTGACGCGGCGTTTGTGCTCTCACACCGACACCTA 67
```

RESULT 6

```
AX244735/c
LOCUS       AX244735                  556 bp    DNA      linear   PAT 28-SEP-2001
DEFINITION  Sequence 64 from Patent WO0166750.
ACCESSION   AX244735
VERSION     AX244735.1   GI:15859614
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Vogeli,G. and Wood,L.S.
  TITLE     G protein-coupled receptors
  JOURNAL   Patent: WO 0166750-A 64 13-SEP-2001;
            PHARMACIA & UPJOHN COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..556
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             34.3%;  Score 378.4;  DB 6;  Length 556;
Best Local Similarity   98.8%;  Pred. No. 5.4e-67;
Matches  403;  Conservative  0;  Mismatches  1;  Indels  4;  Gaps  2;

Qy       697 AGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTCGTGATGCTGGTTGGTGGCCGTG 756
             ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db         556 AGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTCGTGATGCTGGTTGGTGGCCGTG 497

Qy       757 GTGGTCTGCTTCAGCCTCACCACTCAACATCAAGCAGTTCATGGCGAGAGGGATGCTC 816
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         496 GTGGTCTGCTTCAGCCTCACCACTCAACATCAAGCAGTTCATGGCGAGAGGTATGCTC 437

Qy       817 CACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTTACTGCCCTTCAGGCCACCGTGGCC 876
             ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         436 CACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGGCC 377

Qy       877 CTCATGAACATGAACTGTGGCATTACCCCAATCATTTACTTCTTTGCATCCACCCATTAC 936
             |||||||||||||||||||||||||||| |||||||| |||||||||||||||| ||||||
Db         376 CTCATGAACATGAACTGTGGCATTACCCC-ATCATTTACTTTACTTTGCATCCATCCCATTAC 318
```

```
Qy    937 AGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCC 996
          ||||||||||||||||||||||||||||||||||||||  |||||||||||
Db    317 AGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCT--TCCTCCTCCTCCTCC 261

Qy    997 TCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCAGGTATCACCCAGGCCAGAGGTTCT 1056
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 TCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCAGGTATCACCCAGGCCAGAGGTTCT 201

Qy   1057 ATGTTCTTAGCGGAGCACGTGGTCTAACGTGTCATTTGCTTTATGACT 1104
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 ATGTTCTTAGCGGAGCACGTGGTCTAACGTGTCATTTGCTTTATGACT 153
```

```
RESULT 7
AC016189/c
LOCUS       AC016189      151174 bp    DNA    linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016189
VERSION     AC016189.3   GI:9129134
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151174)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-23M9
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 151174)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslaviy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dekrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,K., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6939915.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT     ------------------ Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------ Project Information
            Center project name: L4399
            Center clone name: 23_M_9
            --------------------
            * NOTE: This record contains 152 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *      1    851: contig of 851 bp in length
            *    852    951: gap of 100 bp
            *    952   1804: contig of 853 bp in length
            *   1805   1904: contig of 815 bp in length
            *   1905   2719: gap of 100 bp
            *   2720   2819: contig of 849 bp in length
            *   2820   3668: gap of 100 bp
            *   3669   3768: contig of 889 bp in length
            *   3769   4657: gap of 100 bp
            *   4658   4757: contig of 843 bp in length
            *   4758   5600: gap of 100 bp
            *   5601   5700: contig of 832 bp in length
            *   5701   6532: gap of 100 bp
            *   6533   6632: contig of 849 bp in length
            *   6633   7481: gap of 100 bp
            *   7482   7581: contig of 866 bp in length
            *   7582   8447: gap of 100 bp
            *   8448   8547: contig of 836 bp in length
            *   8548   9383: gap of 100 bp
            *   9384   9483: contig of 859 bp in length
            *   9484  10342: gap of 100 bp
            *  10343  10442: contig of 881 bp in length
            *  10443  11323: gap of 100 bp
            *  11324  11423: contig of 831 bp in length
            *  11424  12254: gap of 100 bp
            *  12255  12354: contig of 934 bp in length
            *  12355  13288: gap of 100 bp
            *  13289  13388: contig of 847 bp in length
            *  13389  14235: gap of 100 bp
            *  14236  14335: contig of 837 bp in length
            *  14336  15172: gap of 100 bp
            *  15173  15272: contig of 858 bp in length
            *  15273  16130: gap of 100 bp
            *  16131  16230: contig of 843 bp in length
            *  16231  17073: gap of 100 bp
            *  17074  17173: contig of 830 bp in length
            *  17174  18003: gap of 100 bp
            *  18004  18103: contig of 836 bp in length
            *  18104  18939: gap of 100 bp
            *  18940  19039: contig of 846 bp in length
            *  19040  19885: gap of 100 bp
            *  19886  19985: contig of 849 bp in length
            *  19986  20834: gap of 100 bp
            *  20835  20934: contig of 827 bp in length
            *  20935  21761: gap of 100 bp
            *  21762  21861: contig of 882 bp in length
            *  21862  22743: gap of 100 bp
            *  22744  22843: contig of 865 bp in length
            *  22844  23708: gap of 100 bp
            *  23709  23808: contig of 868 bp in length
            *  23809  24676: gap of 100 bp
            *  24677  24776: contig of 826 bp in length
            *  24777  25602: gap of 100 bp
            *  25603  25702: contig of 858 bp in length
            *  25703  26560: gap of 100 bp
            *  26561  26660: contig of 844 bp in length
            *  26661  27504: gap of 100 bp
            *  27505  27604: contig of 845 bp in length
            *  27605  28449: gap of 100 bp
            *  28450  28549: contig of 849 bp in length
            *  28550  29398: gap of 100 bp
            *  29399  29498: contig of 859 bp in length
            *  29499  30357: gap of 100 bp
            *  30358  30457: contig of 866 bp in length
            *  30458  31323: gap of 100 bp
            *  31324  31423: contig of 877 bp in length
            *  31424  32300: gap of 100 bp
            *  32301  32400: contig of 854 bp in length
            *  32401  33254: gap of 100 bp
            *  33255  33354: contig of 864 bp in length
            *  33355  34218: gap of 100 bp
            *  34219  34318: contig of 851 bp in length
            *  34319  35169: gap of 100 bp
            *  35170  35269: contig of 856 bp in length
            *  35270  36125: gap of 100 bp
            *  36126  36225: contig of 100 bp in length
```

```
*  36226: contig of 836 bp in length
*  37062: gap of 100 bp
*  37162: contig of 842 bp in length
*  38003: gap of 100 bp
*  38004: contig of 842 bp in length
*  38104: gap of 100 bp
*  38945: contig of 836 bp in length
*  39045: gap of 100 bp
*  39046: contig of 837 bp in length
*  39882: gap of 100 bp
*  39883: contig of 851 bp in length
*  39983: gap of 100 bp
*  40833: contig of 846 bp in length
*  40933: gap of 100 bp
*  40934: contig of 846 bp in length
*  41779: gap of 100 bp
*  41780: contig of 843 bp in length
*  41879: gap of 100 bp
*  41880: contig of 856 bp in length
*  42722: gap of 100 bp
*  42723: contig of 856 bp in length
*  42823: gap of 100 bp
*  43678: contig of 856 bp in length
*  43679: gap of 100 bp
*  43778: contig of 856 bp in length
*  43779: contig of 856 bp in length
*  44634: contig of 855 bp in length
*  44734: gap of 100 bp
*  44735: contig of 853 bp in length
*  45587: gap of 100 bp
*  45588: contig of 855 bp in length
*  46542: gap of 100 bp
*  46543: contig of 846 bp in length
*  46642: gap of 100 bp
*  46643: contig of 846 bp in length
*  47488: gap of 100 bp
*  47489: contig of 860 bp in length
*  47588: gap of 100 bp
*  47589: contig of 860 bp in length
*  48448: contig of 846 bp in length
*  48548: gap of 100 bp
*  48549: gap of 100 bp
*  49450: contig of 902 bp in length
*  49550: gap of 100 bp
*  49551: contig of 833 bp in length
*  50383: gap of 100 bp
*  50384: contig of 870 bp in length
*  50483: gap of 100 bp
*  50484: contig of 870 bp in length
*  51353: gap of 100 bp
*  51354: contig of 853 bp in length
*  51453: gap of 100 bp
*  51454: contig of 853 bp in length
*  52306: gap of 100 bp
*  52307: contig of 846 bp in length
*  52406: gap of 100 bp
*  52407: contig of 846 bp in length
*  53252: gap of 100 bp
*  53253: contig of 872 bp in length
*  53352: gap of 100 bp
*  53353: contig of 872 bp in length
*  54224: contig of 872 bp in length
*  54225: gap of 100 bp
*  54324: contig of 838 bp in length
*  54325: gap of 100 bp
*  55162: gap of 100 bp
*  55163: contig of 824 bp in length
*  55262: gap of 100 bp
*  55263: contig of 824 bp in length
*  56086: gap of 100 bp
*  56087: contig of 840 bp in length
*  56186: gap of 100 bp
*  56187: contig of 840 bp in length
*  57026: gap of 100 bp
*  57027: contig of 899 bp in length
*  57126: gap of 100 bp
*  57127: contig of 899 bp in length
*  58025: gap of 100 bp
*  58026: contig of 879 bp in length
*  58125: gap of 100 bp
*  58126: contig of 879 bp in length
*  59004: contig of 879 bp in length
*  59005: gap of 100 bp
*  59104: contig of 875 bp in length
*  59105: gap of 100 bp
*  59979: contig of 875 bp in length
*  59980: gap of 100 bp
*  60079: contig of 841 bp in length
*  60080: gap of 100 bp
*  60920: contig of 841 bp in length
*  60921: gap of 100 bp
*  61020: contig of 865 bp in length
*  61021: gap of 100 bp
*  61885: contig of 865 bp in length
*  61886: contig of 807 bp in length
*  61985: gap of 100 bp
*  61986: contig of 807 bp in length
*  62792: gap of 100 bp
*  62793: contig of 852 bp in length
*  62892: gap of 100 bp
*  62893: contig of 852 bp in length
*  63744: gap of 100 bp
*  63745: contig of 849 bp in length
*  63844: gap of 100 bp
*  63845: contig of 849 bp in length
*  64693: gap of 100 bp
*  64694: contig of 852 bp in length
*  64793: gap of 100 bp
*  64794: contig of 852 bp in length
*  65645: gap of 100 bp
*  65646: contig of 879 bp in length
*  65745: gap of 100 bp
*  65746: contig of 879 bp in length
*  66624: contig of 879 bp in length
*  66625: gap of 100 bp
*  66724: contig of 826 bp in length
*  66725: gap of 100 bp
*  67550: contig of 826 bp in length
*  67551: gap of 100 bp
*  67650: contig of 852 bp in length
*  67651: gap of 100 bp
*  68502: contig of 852 bp in length
*  68503: gap of 100 bp
*  68603: contig of 833 bp in length
*  69435: contig of 833 bp in length
```

```
Qy     681  GGAGGACCCAGTGACCAGCAGAAAGAC-GCCACTGGCGAGGCTGCTGCCTTACGGTGC    739
            |||||     |||||     |||||| ||| ||||||||||||||||||||||||||||
Db  107428  GTAGCATTCAGTTCCAGGAGTAAAGCACCGCCTCTGCCGAGGCTGCTGCCTTACGGTGC  107369

Qy     740  TGATGCTGGTGGCCGTGGTGGTCTGCTTCAGCGCAGTTC                       798
            ||||||||||||||||||||||   || |||||||          ||||||||||
Db  107368  TGATGCTGGTGGCCGTGGTGGTTTGTTTCAGCGCCCTCAACATCAAGCAGTTC       107309

Qy     799  ATGGCGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGCGGGCTTTCTTACTGTCC   858
            ||||     |||||||||||||||||||||||||||||||||  |||||||||||||||
Db  107308  ATGGTGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGAGGGCTTTCTTACTGTCC  107249

Qy     859  CTTCAGGCCACCGTGGCCCTCATGAACATGAACTGTGGCATTACCCCAATCATTTACTTC  918
            ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  107248  CTTCAGGCCACCGTGGCCCTCATGAACATGAACTGTGGCATTACCCC-ATCATTTACTTC  107190

Qy     919  TTTGCATCCACCATTACAGGAAATGGCTCCTGGCCACCCAGAAAGCTTCAAAGGGTCTTC  978
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  107189  TTTGCATCCACCATTACAGGAAATGGCTCCTGGGCCACCCAGAAAAGCTTCAAAGGGTCTTCC  107130

Qy     979  TCCTCCTCCTCCTCCTCCTCCTCCCCAGAAAAGTTCTTCAGAAACACCAAGTATC      1038
            |||||||||||||||||||||||  ||||||  ||||||||||||||||||||||
Db  107129  TCCTCCTCCTCCTCCTCCTCCACCCCAGAAAAGCTTCTTCAGAAACACCAAGTATC    107070

Qy    1039  ACCCAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTG-GTCTAACGTGTCATTTGCTT 1097
            |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db  107069  ACCCAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTGCGTCTAACGTGTCATTTGCTT  107010

Qy    1098  TATGACT  1104
            |||||||
Db  107009  TATGACT  107003
```

```
RESULT 8
BX470183/c
LOCUS       BX470183              162543 bp    DNA     linear   HTG 04-AUG-2003
DEFINITION  Danio rerio clone CH211-184M13, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   BX470183
VERSION     BX470183.3  GI:33343489
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162543)
AUTHORS     McLaren,S.
TITLE       Direct Submission
JOURNAL     Submitted (03-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 4, 2003 this sequence version replaced gi:30424174.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zC184M13
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 161980 bases at least Q40
            Consensus quality: 162049 bases at least Q30
            Consensus quality: 162109 bases at least Q20
            Insert size: 162243; sum-of-contigs
            Insert size: 177307; 8.8% error; agarose-fp
            Quality coverage: 8.74x in Q20 bases; sum-of-contigs Quality
            coverage: 8.00x in Q20 bases; agarose-fp
            * NOTE: This is a 'working draft' sequence. It currently
```

```
Query Match    32.0%;   Score 353.4;   DB 2;   Length 151174;
Best Local Similarity   95.3%;   Pred. No. 8.2e-62;
Matches 407;   Conservative 0;   Mismatches 16;   Indels 4;   Gaps 4;
```

```
Db  133070  GACCTGGTCATCGAACGTGCATGAGAGTTCTCCAATTTGACAG--CAAGAGGATGGCCT  133014
                |||||||||||||| ||||||||||| ||||||||| ||||||     |||||||||||
Qy     581  TCATGGTCCTGGTGGCCTTTGCCATTGGCCTTCTGTGGGCCAGTGGGGATCATCGTGTCCT  640

Db  133013  ATGTTCTTCTGTTCGTGCTTGCGTCATTGGGTTTTTCATGCCATTGGGACTTATTCTTTGCT  132954
                ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Qy     641  GCTATATGAAGATCACCTGGAAGCTGTGCAGCAGCGGCACAGCTCCGGAGGACCCAGTGACCAGCA  700

Db  132953  CCTACAGCCGGGTCAGCTGCAAGCTCTACAGACGGGCCAAGAAAACCCTTTGACCAACA  132894
                |||||||||  ||||||||||||||||||  |||||||  |||||||||| || |||
Qy     701  GGAAAGGACGGCCATGGCGAGGCTGCCTGCTTACGCTGCTGATGCTGGTGGCCGGTGGTGG  760

Db  132893  AGTCTGGAGGCCACACCAGAGCCAAAAATAATGATCTTACTCATTCGTCCAGTTTCGTAG  132834
                ||||||||||| ||| |||||||||||||||||| |||||||||| || |||||||||
Qy     761  TCTGCTTCAGCCGCCACCACCTCAACATCACAGTTCATGGCGAGAGGATGCTCCACC  820

Db  132833  TGTGCTTCAGCCCTTACCACATTACATCACATGCAGTTTGTTGTGAGAAGGCTTTACCACG  132774
                | ||||||||| |||||||||||||| ||||| |||||||||||||||||||||||||
Qy     821  TGCCATCCTGTGCCGAGCGGAGGGCTTCTTACTGTCCCTTCAGGCCACCGTGGCCCCTCA  880

Db  132773  AGCCCCTCTTGTGAAGAGTTAAGGCTTTGAAATTGTCTCTGCAGTAACCGTTGCCATGA  132714
                |||||||||||| ||||||| ||||| |||||||||||||||||||||||||||||||
Qy     881  TGAACATGAACTGTGCATTACCCCAATCATTACTTCTTTGCATCACCATTACAGGA  940

Db  132713  TGAACTTAAACAGTGCGCTTGCCTTGACCCTATTATTACTTTTTTGCCATAGAACGTATAAAGC  132654
                |||||||| |||||||||||||||||||||  |||||| |||||||||||||||||||
Qy     941  AATGGCTCCTGGCATTTAAAGCTCAAAGGGTCTTCCTCCCTCCTTCCTCC  990

Db  132653  AGAGAGTCATGAGTCCTTTTAAGAGCTACATTTCCATTCGGCAGCCTCC  132604

RESULT 9
AC102718
LOCUS       AC102718          190588 bp    DNA     linear   HTG 14-JUN-2003
DEFINITION  Mus musculus clone RP24-469E6, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC102718
VERSION     AC102718.3  GI:31745314
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 190588)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-469E6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190588)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Dias,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,J., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,D., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
    * consists of 4 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *
    *        1      5005: contig of 5005 bp in length
    *     5006      5105: gap of 100 bp
    *     5106    101540: contig of 96435 bp in length
    *   101541    101640: gap of 100 bp
    *   101641    156970: contig of 55330 bp in length
    *   156971    157070: gap of 100 bp
    *   157071    162543: contig of 5473 bp in length.

FEATURES             Location/Qualifiers
     source          1..162543
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-184M13"
                     /clone_lib="CHORI-211"
     misc_feature    1..5005
                     /note="assembly_fragment:01046
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5106..101540
                     /note="assembly_fragment:01887
                     fragment_chain:1"
     misc_feature    101641..156970
                     /note="assembly_fragment:00490
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
     misc_feature    157071..162543
                     /note="assembly_fragment:01214
                     fragment_chain:2"

ORIGIN

Query Match             23.7%;   Score 261.2;   Pred. No. 5.8e-43;   DB 2;   Length 162543;
Best Local Similarity   56.6%;
Matches  504;   Conservative   0;   Mismatches 383;   Indels  3;   Gaps  1;

Qy     101  CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC  160
                |||||||||||||  ||||| |||||||||||| || ||||||||||||||||| ||
Db  133490  CACTTTTCTACTGCCTGATTTTTGTGATCAGCGTGGCAGGTAACTCCCTGGGTCCTCTGCA  133431

Qy     161  TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG  220
                ||||||| ||||| ||| |||| |||||| |||||||||  ||||||||||||||||
Db  133430  TCACCTGTCAGAAAAGGCAGAAGATGAACTTCTACCACATATATTGATCAACCTGGCCA  133371

Qy     221  TGTCTGACCTGCTGTTCACCGTGGCCTTACCCGGAAGGGTGGTGTGTATGTGCTGGGCT  280
                ||||| ||||||||  ||||||||||||||| |||||||||| |||||||||||||
Db  133370  TCTCCGATGCTCTCTTCACATTAGCTGTCTGCCCGGCAGGATCACCTACTACCCGTGGCT  133311

Qy     281  CCAGTGCGCCTTCGGCAAGGGGGCTCTGCAAGGGGGCTGACGGCGTTTGTGTCTCTACACCGACA  340
                ||||||||||||||||||||  ||||||||  ||||| ||||||||||||||||||||
Db  133310  TTGATTGGCCTTTGGAGATTCTTGTGCCGCCGTGAGCGGTGAGATCTTCTACAGCAATA  133251

Qy     341  CCTACGGGGGGTCTACCTCATGGCCGTGTGTGAGCGGTGACCATTACCCAGCTGTGGTCT  400
                |||||||||||||||| ||||||| ||||||||||| ||| ||||||||||||| |||
Db  133250  CCTATGTAAGCACATCGCTTTCATGACCTGCATCAGTGTGGATCGGCTACTTGGCCATGCTGC  133191

Qy     401  GTGCCCACTGGGGCCCCGGCCTCCGCACGGCGTGGCTCGCCCCAGGCGTCTGGCGGCCA  460
                |||||||||||||||||||||||||  ||||||||||||||||| ||||||||||||
Db  133190  ACCGTCAGCGCTGTCAAAGACTGAGAAAAACCAAAGTCGTGAAAGCCGGTCGTGTTCTGG  133131

Qy     461  TCTGGACCTGGTGCTGCGCAGAGAGATGCCCTTGCTTCCTTGATGCCAGTGACCAAGCCGC  520
                || ||||||| || ||||||||||||||  |||||||||||||| ||||||||||||
Db  133130  TGTGGGTGGTGGTTCGTTCCTTCATGAAACATCACCATTGCTTTCAAGAGCATCATGGAAAACA  133071

Qy     521  TGGGTGGGCAAGCTGCCTGACGTAGTAGCAGCAGCATGGAGTCAGTCCTCGGGCTGCCCC  580
                ||||||||||||||||||||||||||| ||||||||||||| ||| ||
```

```
REFERENCE    3 (bases 1 to 190588)
AUTHORS      Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
             Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
             Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
             Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
             Diaz,J.S., Dodge,S., Dooley,K., Dorris,J., Erickson,J., Faro,S.,
             Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
             Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
             Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
             Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
             Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
             Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
             Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
             O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
             Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
             Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
             Spence,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
             Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
             Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
             Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (14-JUN-2003) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Jun 14, 2003 this sequence version replaced gi:22381711.
             All repeats were identified using RepeatMasker:
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             --------------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence submissions@genome.wi.mit.edu
             ---------------- Project Information
             Center project name: L19343
             Center clone name: 469_E6
             ---------------- Summary Statistics
             Sequencing vector: Plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 188732 bases at least Q40
             Consensus quality: 188884 bases at least Q30
             Consensus quality: 189429 bases at least Q20
             Insert size: 183000; agarose-fp
             Insert size: 189788; sum-of-contigs
             Quality coverage: 10.2 in Q20 bases; agarose-fp
             Quality coverage: 9.9 in Q20 bases; sum-of-contigs
             --------------------
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 9 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record as it is available and the accession number will
           * as soon as it is available and the accession number will
           * be preserved.
           *        1      330:  contig of 330 bp in length
           *      331      430:  gap of 100 bp
           *      431      970:  contig of 540 bp in length
           *      971     1070:  gap of 100 bp
           *     1071     7799:  contig of 6729 bp in length
           *     7800     7899:  gap of 100 bp
           *     7900    16142:  contig of 8243 bp in length
           *    16143    16242:  gap of 100 bp
           *    16243    29388:  contig of 13146 bp in length
           *    29389    29488:  gap of 100 bp
           *    29489    49346:  contig of 19858 bp in length
           *    49347    49446:  gap of 100 bp
           *    49447   101875:  contig of 52429 bp in length
           *   101876   101975:  gap of 100 bp
           *   101976   136840:  contig of 34865 bp in length
           *   136841   136940:  gap of 100 bp
           *   136941   190588:  contig of 53648 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..190588
                     /organism="Mus musculus"
                     /mol type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-469E6"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1..330
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    431..970
                     /note="assembly_fragment"
     misc_feature    1071..7799
                     /note="assembly_fragment"
     misc_feature    7900..16142
                     /note="assembly_fragment"
     misc_feature    16243..29388
                     /note="assembly_fragment"
     misc_feature    29489..49346
                     /note="assembly_fragment"
     misc_feature    49447..101875
                     /note="assembly_fragment"
     misc_feature    101976..136840
                     /note="assembly_fragment"
     misc_feature    136941..190588
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Query Match    20.1%;  Score 221.6;  DB 2;  Length 190588;
Best Local Similarity  53.1%;  Pred. No. 7.3e-35;
Matches 498;   Conservative   0;  Mismatches 434;  Indels   6;  Gaps   1;

Qy      101 CTCTGTTCTACACAGCCTCCTCGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
                ||||| ||||||||||||| |||||||||||||| ||||||||  |||||  ||||||
Db   107260 CTCTGCATTACAGCCTGGTCTTCATCATTGGGCTGGTGGGAAACCTGCCTGGCCTTGGTTG 107319

Qy      161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
                |||||| ||||  ||| |||| || |||| || ||||  |||||||||||||||
Db   107320 TCATTGTTCAAAACAGAAAAAAAATCAACTCAACCACTCCTATTCAATGAACTTGGTCA 107379

Qy      221 TGTCTGACCTGCTGTTCACCGTGCCTACCGGGAAGGGTGGTGTGTTATGTGCTGGGCT 280
             ||||| ||||| || ||||||  ||| |||| ||||||||||||||||||||||||||
Db   107380 TTTCTGACATCCTGTTTACCACAGCTTTACCCACTCGGATAGCCTACTATGCGCTGGGCT 107439

Qy      281 CCAGCTGGCCTTTCGGCAAGGGCTCTGCAAGGGCTCGCAGGCTGACGGCGTTTGTGCTTCTACCGACA 340
             |||||||||||||||||| |||||| ||||| |||||||| ||||| |||  || |||||||||||||
Db   107440 TTGATTGGAGGATAGGTGATGCCCTGCCGTGCCGGGTAACTGCTCCTGGTGTTCTCACATCAACA 107499

Qy      341 CCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCTGAGTGTGACGCGTGGGTCT 400
                |||||| |||||||||||  ||||||||||| |||| |||||| ||||| ||||||||
Db   107500 CGTACGCAGGTGTGAACTTCATGACTTGCTTGAGCATAGACCGCTTCTTCGCGTGGTGC 107559

Qy      401 GTGCCCACTGGGGCCGCGGCCTCCGCACGGCTGCCGCGGCCAGGCTGGTCTGCGTGGCCA 460
             ||||||||| |||||||||||||||||||  || |||| ||||||||||||||||||||||
Db   107560 ACCCTCTGCCGCTACACACAAGATTAAAGAAATCGAATACGCAAAGGGTGTCTGCCTGTCCG 107619

Qy      461 TCTGGACCTGGTGCTGCTGCGACGATGCCCTTGCTCTGATGCCCATGACCAAGCCGCC 520
                ||||| |||||| |||| |   ||||| ||| ||||||||| |||||||  |||||
Db   107620 TCTGGATTCGGTCTTTGCTCAAACACTGCGTCTCCTCACCCTATGTCTAAGGAGG 107679

Qy      521 TGGTGGGCAAGCTGCCTGCATGGAGTACAGCAGCATGGACTGCAGTCGTTGCCCC 580
                ||||||||||| ||||||||||||| ||||||||  ||||||| ||||| |||||
Db   107680 AGGGAGACAAGACCACTTGCATGGAGTATCCAAACTTTGAAGGGACAGCGTCCCCTGCCGT 107739

Qy      581 TCATGGTCCTGGTGGCCTTTGCCATTGGGCCATTGCCTTCGTGGGCCAGTGGGGATCATCCTGTCCT 640
             ||||| |||||||||||||||| | ||||||||||||| ||||||  ||||||||||||||||||
Db   107740 GGATTCTGCTCCGGAGCCTGTCTGCTGGGCTAGCGTGCTCCTATCACAGTCATTCCCTGT 107799

Qy      641 GCTATATGAGATACACCTGGAAGCTGGACAGTGTGCAGCACAGCTCCGGAGGACCCAGTGACCAGCA 700
                ||||  ||||  |||||| |||| ||||||||
```

```
Db  107800 GTTACTCTCGAGATCTGCTGCAAACTCTTCAGGACTGCCAAGCAGAGAACCCACTCACCGAGA 107859
Qy     701 GGAAAGGACGGCCACTGGGCAGGAGGCTGCGCTTACGCTGCTGATGCTGGTGGCCGGTGGTGG 760
Db  107860 AATCTGGTGTGAGACAAAAGGCTCTCAACACAATTATCCTCATCATTGCGTGTTCATCC 107919
Qy     761 TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGCGAGAGGATGCTCCACC 820
Db  107920 TGTGCTTCACGCCCTACCACGTGGCCATCATTCAGCACATGACATGAAGATGCTCTGCTCCC 107979
Qy     821 TGCCATCC------TGTGCCAGCGGAGCGGAGGGCTTCTTACTGTCCCTTCAGGCCACCCTGG 874
Db  107980 CTGGAGCCCTGGAGTGTGGGCGGAGACAATTCCTTCCAGATCTCCTGCACTTCACGGTGT 108039
Qy     875 CCCTCATGAACATGACATGTGGCATTACCCCAATCATTTACTTCTTTGCATCCACCCATT 934
Db  108040 GCCTGATGAACTTCAACTGCTGCATGGACCCGGTTCATATACTTCTTTGCATGCAAAGGGGT 108099
Qy     935 ACAGGAAATGCTCCTGGGCATTTTAAAGTCAAAGGGGTCTTCCTCCTCCTCCTCCT 994
Db  108100 ATAAGAGAAAGGTCATGAAGATGCTCAAACGTCAAGTGAGTGGTGTCATATCCAGCCAG 108159
Qy     995 CCTCCTCCACCCAGGAAAAGCTTCTTCAGAAACACCA 1032
Db  108160 TGAGGTCAGCCCCTGAAGAGAGAATTCGCGGGGAAATGACA 108197
```

RESULT 10
AC110249
AC110249
LOCUS   AC110249   204438 bp   DNA   linear   HTG 25-NOV-2003
DEFINITION Mus musculus chromosome 14 clone RP23-25708 map 14, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC110249
VERSION   AC110249.9 GI:38502441
KEYWORDS   HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 204438)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Mus musculus chromosome 14, clone RP23-25708
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 204438)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Grand-Pierre,N.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwar,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,T., Nguyen,J., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,C.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramarian,A., Talamas,J., Tesfaye,J., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 204438)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,

            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,J.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbit,R., Maclean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,v.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Nov 25, 2003 this sequence version replaced gi:37777380.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ------------------------- Project Information
              Center project name: L22560
              Center clone name: 257_O_8
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1 168488: contig of 168488 bp in length.
            * 168489 168588: gap of 100 bp
            * 168589 204438: contig of 35850 bp in length.
FEATURES   Location/Qualifiers
     source   1..204438
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="14"
                     /map="14"
                     /clone="RP23-25708"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

Query Match   20.1%; Score 221.6; DB 2; Length 204438;
Best Local Similarity 53.1%; Pred. No. 7.3e-35;
Matches 498; Conservative 0; Mismatches 434; Indels 6; Gaps 1;

Qy   101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTCGGGAAACATCCTTGCCCTTTGCC 160
           ||||| |||| ||| ||||| | || ||| ||| ||||||| |||| | |||
Db 160134 CTCTGCATTACAGCCTGGTCTTCATCATTGGGCTGGTGGGAAACCTGCTGCTGGCCTTGTGGTTG 160193
Qy   161 TTACCTGTCAAAAGACAGGAAGATCAACTGACAGGCATCAACAGCCATCACCTGGCTG 220
           |||| || ||||| | ||| ||||||| | || | |||||| | |||| ||
Db 160194 TCATTGTTCAAAACAGAAAAAAATCAACTCAACCACTCTCATTCAATGAACTTGGTCA 160253
Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGAAGGGTGGTGTTATGTGTCTGGGCT 280
           ||||| ||||| |||| || | | |||| || |||||| | | | ||||||
Db 160254 TTTCTGACATCCTGTTTACCACAGCTTTACCCACTCGGATAGCCTACTATGCGCTGGGCT 160313
Qy   281 CCAGCTGGCCTTTCGGCAAGGGGCCTGCGAGGCTGACGCCGTTTGTGCTCTACACCGACA 340
```

Db 160314 TTGATTGGGAGGATAGGTGATGGCCCTGTGCTGCGGGTAACTGCTTCTTCGTGTTCTACATCAACA 160373
           |||||  || ||| ||| ||||| |||||||||||  |||||||||||  || || |||||||||||
Qy   341 CCTACGGGGGGGTCACCTGGGTCGTGTGAGCGTGACCAATTACCCAGCTGTGGTCT 400

Db 160374 CGTACGCAGGTGTGAACTTCATGACTTGCTTGAGCATAGACCCGCTTCTTCGCTGTGGTGC 160433
           || || |||||||||| ||| ||||| || || |||||| ||| |||| |||||||||||
Qy   401 GTGCCCACTGGGGCCCGGGCGCCTGCCGCACGGCCAGGCTGGTGCTGCGTGCGCCA 460

Db 160434 ACCCCTGCGCTACACACAAGATTAAAGAATCGAATACGCAAAGGGTGTTGCCTGTCCG 160493
           |||||||||| || || |||||||||||||| |||| |||| |||| || ||||||||
Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATCCCATGACCAAGCCGC 520

Db 160494 TCTGGATTCGGTCTTGCTCAAACACTGCGCTGCTCCTCACCCTATGTCTAAGGAGAGG 160553
           ||||| || || |  || || || ||||  || |||||||||  ||  ||||||||||
Qy   521 TGGTGGCAAGCTGCCTGCGATGGAGTACAGCAGCCATGGAGTCAGTCCTCGGGCTGCCCC 580

Db 160554 AGGGAGACAAGACCACTTGCACTTGCATGGGAGTATCCAAACTTTGAAGGGACAGCGGTCCCTGCCGT 160613
           ||||||||||||  ||||| |||||||||||||||  || || || ||||||||| || ||||||||
Qy   581 TCATGGTCGTGGCCTTTGCCATTGGCTTCTGTGGGCCCAGTGGGATCATCCTGTCCT 640

Db 160614 GGATTCTGCTCGGAGCTGTGTCTGCTGGGCTACGTGCGTCGCTGCTCTATCACAGTGCATTCTCCTGT 160673
           ||  |||| || ||| || || |||||||| || ||| |||| || || ||| ||| ||||||||||
Qy   641 GCTATATGAAGATCACCTGGAAGCGTGCAGCACAGTCGGGAGGACCCAGTGACCAGCA 700

Db 160674 GTTACTCTCAGATCTGCTGCAAACTCTTCAGGACTGCCAAGCAGAACCCACTCACCGAGA 160733
           ||||||||| |||||| |||| ||| ||| ||||| ||||| || ||| || || |||||
Qy   701 GGAAAGGACGGCCACTGGCGAGGCGTGCCTGCTTACGCTGCTGATGCTGGCCGTGGTGG 760

Db 160734 AATCTGGTGTGAACAAAAAAGGCTCTCAACACAATTATCCTCATCATTGTGTGTTCATCC 160793
           || ||||||||||| ||||||| || ||  || || ||| || || || || ||||||||
Qy   761 CCCTCATGATGAACAATGGCTTCATGGCCGAGCAGTTCATGGCCGAGAGGGATGGCTCCACC 820

Db 160794 TCTGCTTCAGCCCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATGGCCGAGAGGGATGGCTCCACC 160853
           ||||||||||| || ||||| ||||||||||||||||||||||||||||||||||||||||||
Qy   821 TGTGCTTCACGCCCTTCACCACCACGGTTCAGGCCACCACCGTGG 874

Db 160854 CTGGAGCCCTGGAGTGTGGGGGCGAGACATTCCTTCCAGATCTCTCTCCAGTCTCTCAGCCTCACGGTGT 160913
           || ||||||||| |||| ||||| ||| || |||| |||||| ||||  |||||||| || || ||||||
Qy   875 CCCTCATGAACAATGGCTTCATGGCCGAGCAGTTCATGGCCGAGAGGGATGGCTCCACC 934

Db 160914 GCCTGATGAACTTCACTGCTGCATGGACCCGTTCATGATACTTCTTTGCATGCAAAGGGT 160973
           |||| ||||||| ||  || ||||| ||| ||||||| |  || ||| ||||| ||||||
Qy   935 ACAGGAAATGGCTCCTCGGCACGAT TTTAAAGCTCAAAGGGGTTCTTCCCTCCTCCTCCCTCCT 994

Db 160974 ATAAGAGAAGGTCATGAAAGTCAAACGTCAAGTGAGTGTGTCGATTCTCAGCGCAG 161033
           || ||||||||| |||||||| ||| ||||| ||||| || || |||||| ||||||
Qy   995 CCTCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCA 1032

Db 161034 TGAGGTTCAGCCCCCTGAAGAGAGAATTCGCGGGAAATGACA 161071
           ||||||||||||  |||||||||||||| ||||||||||||||
Qy  1033 TGAGGTTCAGCCCCCTGAAGAGAGAATTCGCGGGAAATGACA 1074

RESULT 11
BC052868
LOCUS         BC052868          1493 bp    mRNA        linear    ROD 08-OCT-2003
DEFINITION    Mus musculus Epstein-Barr virus induced gene 2, mRNA (cDNA clone
              MGC:60769 IMAGE:30073287), complete cds.
ACCESSION     BC052868
VERSION       BC052868.1    GI:31112264
KEYWORDS      MGC.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 1493)
AUTHORS       Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
              Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
              Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
              Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
              Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
              Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
              Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
              Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,

Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE         Generation and initial analysis of more than 15,000 full-length
              human and mouse cDNA sequences
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE       22388257
PUBMED        12477932
REFERENCE     2 (bases 1 to 1493)
AUTHORS       Strausberg,R.
TITLE         Direct Submission
JOURNAL       Submitted (27-MAY-2003) National Institutes of Health, Mammalian
              Gene Collection (MGC), Cancer Genomics Office, National Cancer
              Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
              USA
REMARK        NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT       Contact: MGC help desk
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Drs. Dennis Taub, Dan Longo (NIA, USA),
              Jonathan Keller (NCI, USA)
              cDNA Library Preparation: Yulan Piao and Minoru Ko (National
              Institute on Aging, NIH: http://lgsun.grc.nia.nih.gov/cDNA/)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: National Institutes of Health Intramural
              Sequencing Center (NISC),
              Gaithersburg, Maryland;
              Web site: http://www.nisc.nih.gov
              Contact: nisc_mgc@nhgri.nih.gov
              Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
              Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
              Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
              Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
              Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
              McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
              Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
              Young,A., Zhang,L.-H. and Green,E.D.

              Clone distribution: MGC clone distribution information can be found
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              Series: IRAK Plate: 112 Row: o Column: 14
              This clone was selected for full length sequencing because it
              passed the following selection criteria: Hexamer frequency ORF
              analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1493
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6NCr"
                     /db_xref="taxon:10090"
                     /clone="MGC:60769 IMAGE:30073287"
                     /tissue_type="Hematopoietic Stem Cell,
                     (Lin-/c-Kit-/Sca-1+), mouse, 10 wks"
                     /clone_lib="NIA Mouse Hematopoietic Stem Cell
                     (Lin-/c-Kit-/Sca-1+) cDNA Library (Long)"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1"
     gene            1..1493
                     /gene="Ebi2"
                     /db_xref="LocusID:321019"
                     /db_xref="MGI:2442034"
     CDS             73..1146
                     /codon_start=1
                     /product="Epstein-Barr virus induced gene 2"
                     /protein_id="AAH52868.1"
                     /db_xref="GI:31127265"
                     /db_xref="LocusID:321019"

```
          /translation="MANNFTTPLATSHGNNCDLYAHHSTARVLMPLHYSLVFIIGLVG
NLLALVIVQNRKKINSTTLYSMNLVISDLFFTALPTRIAYYALGFPWRIGDALCRV
TALVFVINTYAGVNFMTCLSIDRFPAVVHPLRYNKIKRIEYAKGVCLSWILVFAQTL
PLFLTPMSKEEGDKTTCMEYPNFEGTASLPWILLGACLLGYVLPITVILLCYSQICCK
LFRTAKQNPLTEKSGVNKKALNTILIIIVVFFILCFTPYHVAIICHMIMLCSPGALEC
GARHSFQISLHFTVCLMNFNCCMDPPIYFFACKGYKRKVMKMLKRQVSVSISSAVRSA
PEESSREMTESQMMIHSKASNGR"

misc_feature   202..984
          /note="7tm_1; Region: 7 transmembrane receptor (rhodopsin
          family)"
          /db_xref="CDD:pfam00001"

ORIGIN

Query Match         19.9%;   Score 220;  DB 10;  Length 1493;
Best Local Similarity   53.0%;
Matches 497;  Conservative    0;  Mismatches 435;   Indels    6;   Gaps    1;
```

```
Qy   101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
         ||||||    ||| |  ||||||  ||||     |||| ||||| ||  ||||  || |||
Db   164 CTCTGCATTACAGCCTGTCTTCATCATTGGGCTGTGGAAACCTGCTGGCCTTGGTTG 223

Qy   161 TTACCTGTCAAAAGACGAGAAGATCAACTCAACTCAACCACTCTCTATTCAATGAACTTGGTCA 220
         ||||  | ||||| | |   | | | || ||| ||  | |||||| |  |  ||| ||| ||
Db   224 TCATTGTTCAAAACAGAGAAAAAAATCAACTCAACCACTCTCTATTCAATGAACTTGGTCA 283

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCCGGAGGGTGGTGTTGTTATGTGCTGGGGCT 280
         | | |||| |||| | | |  |  ||| || ||| |  |||| ||| ||| | || |||||
Db   284 TTCTGACATCCTGTTTACCAGCGTTTACCCACTCCGGATAGCCTCGATGCTCGCTCGGGCT 343

Qy   281 CCAGCTGGCCTTTCGGACAAGGGGCTCTGCAGGCTGACGGCGTTGTGCTCTCACACCGACA 340
         |||| ||||| ||  | | ||  |  | |  ||| ||| |  | || || |  |||| ||
Db   344 TTGATTGGAGGATAGGTGATGCCCGTGCCCGGGTAACTGCTCCTGGTGTTCTACATCAACA 403

Qy   341 CCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCCTGGACCATTACCCAGCTGTGGTCT 400
         ||||||  |  |  | |  |||| ||  ||| ||||| || |||||  ||||||||||||
Db   404 CGTACGCAGTGTGAACTTCATGACTTGCCTCTCGATAGACCCGCTTCCGCTGTGGTGC 463

Qy   401 GTGCCCACTGGGGCCGCGCCTCCGCACGGCGTGGCTGGGCCCAGGCGTGGTCGTGGTGGCCA 460
         ||||||| | |||||  |||| |  | | | ||  |  ||| |||  |||||| ||| ||||
Db   464 ACCCTCTGCGCTCAACAAGATTAAAGAATGAATACGCAAAGGGTGTCTGCCTGTCCG 523

Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCCTTGATGCCCATGACCAAGCCGC 520
         |||| || |||| | |||| | |||| | |  |||| |||||  ||||| |||| || ||
Db   524 TCTGATTCTGGTCTTGCTCAAAACACTGCCCGTGTTCCTCACCCCTATGTCTAGGAGG 583

Qy   521 TGGTGGGCAAGCTGGCCTGCATGGAGTACCAGCAGCAGCATGGAGTCAGTCCTCGGGGCTGCCCC 580
         | | || ||||| || |  ||||||| || |||| | ||  |   ||||  | | ||| |||||
Db   584 AGGGAGACAAGACCACTTGCATGGAGTATCCAAACTTTGAAGGGACAGCGTCCCTGCCGGT 643

Qy   581 TCATGGTCCTCGGTGCCTTTGCCATTGGCTTCTGTGGGCCCAGTGGGGATCATCCTCGTCCT 640
         |||  | | ||  | ||  |||||||| |||||  |  ||| |  || || || | ||
Db   644 GGATTCTGCTCGGAGCCCTGTCTGCTGGTACGTGCCTACCGTGCCTATCACAGTCATTCCTCTGT 703

Qy   641 GCTATATGAAGATCACCTGGAAGCTGTGCAGACAGCTCGGAGGACCCAGTGACCAGCA 700
         ||| |  |||| | |||| |||   | || |  | | || |||||||||   ||||||
Db   704 GTTACTCTCAGATCTGCTGCAAACTCTTCAGACTGCCAAGCAGAACCCACTGACCACGAGA 763

Qy   701 GGAAAGGACGCCCACTGGGCAGAGGCTGCCCTCTACGCGCTGATGCTGGTCGGTGGTGG 760
         ||| || ||  ||  |||  | |||||| | ||    ||| | | || || | ||||||
Db   764 AATCTGGTGAACAAAAAGGCTCTCAACACCATTATCCTCACCATTATTGTCGTTTCATCC 823

Qy   761 TCTGCTTCAGCCCCTGACCCTACCACCTCAAACATCAAGCAGTTCATGCCGAGAGGAGAGCTCCACC 820
         |||||| ||  ||| |  || | ||| |||| ||| | | ||| |    |    | ||||
Db   824 TGTGCTTCACGCCCTACCACGTGGCCATCATCTGCCACATGATCATGATACTCTGCTCCC 883

Qy   821 TGCCATCC------TGTGCCGAGCGGAGGCTTTCTTACTGTCCCCTTACGCACCGCACCGTGG 874
         ||||            |  |||||  |  | | | |  ||||  |  |  | | |  ||  |
Db   884 CTGGAGCCCTGGAGTGTGGGGCCGAGAGACATTCCTTCCAGATCTCTCTGCACTTCACGGTGT 943

Qy   875 CCCTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
         |  ||||  ||   |  |  ||| | || |    ||| |   || || ||  | | |||
Db   944 GCCTGATGAACTTCAACTGCTGCATGGACCCGCCTATCTATTTCTTTGCATGCAAAGGGT 1003
```

```
Qy   935 ACAGGAAATGGCTCCTGGCATTTAAAGCTCAAAGGGTTCTTCCTCTCCTCCTCCTCCT 994
         || |||||| | |  |||  |||   | ||| ||| ||||||||| || || || ||
Db  1004 ATAAGAGAAAGGTCATGAAGATGCTCAAACGTCAAGTGAGTGTGTCGATCTCCAGCGCAG 1063

Qy   995 CCTCCTCACCCCAGGAAAAGCTTCTCAGAAACACCA 1032
         || || |||  |  ||   |   |||  ||||| |||
Db  1032 TGAGGTCAGCCCCTGAGAGAGAGTTCGGCGGAAATGACA 1101
```

```
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 239344)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (10-SEP-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3  (bases 1 to 239344)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Nov 19, 2002 this sequence version replaced gi:22771237.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
             ------------ Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
             ------------ Project Information
                Center project name: KANY
                Center clone name: CH230-64N15
             ------------ Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 218763 bases at least Q40
                Consensus quality: 220754 bases at least Q30
                Consensus quality: 221894 bases at least Q20
                Estimated insert size: 217944; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
             ----------
             * NOTE: Estimated insert size may differ from sequence length
             *     (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             *     consists of 3 contigs. The true order of the pieces
             *     is not known and their order in this sequence record is
             *     arbitrary. Gaps between the contigs are represented as
             *     runs of N, but the exact sizes of the gaps are unknown.
             *     This record will be updated with the finished sequence
             *     as soon as it is available and the accession number will
             *     be preserved.
             *         1     17366: contig of 17366 bp in length
             *     17367     17466: gap of unknown length
             *     17467    228779: contig of 211313 bp in length
             *    228780    228879: gap of unknown length
             *    228880    239344: contig of 10465 bp in length
FEATURES             Location/Qualifiers
     source          1..239344
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-64N15"
     misc_feature    complement(12152..13760)
                     /note="clone_boundary
                     clone_end:Sp6"
                     site:
                     end_sequence:BH346389"
     misc_feature    13069..14398
                     /note="wgs_contig"
     misc_feature    14449..15935
                     /note="wgs_contig"
     misc_feature    16002..17366
```

```
     misc_feature    /note="wgs_contig"
                     complement(209286..210084)
                     /note="clone_boundary
                     clone_end:T7"
                     site:
                     end_sequence:BH346386"
     misc_feature    228880..229915
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    234188..235777
                     /note="wgs_end_extension
                     clone_end:T7"

ORIGIN

Query Match       18.9%;  Score 208.6;  DB 2;  Length 239344;
Best Local Similarity   53.0%;  Pred. No. 3.4e-32;
Matches 4473;  Conservative  0;  Mismatches 414;  Indels  6;  Gaps  1;

Qy     101 CCTCTGTTCTACACAGCCCTCCTGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
           |||||| |||| |||  |  |||||| || || |||||||||||||| |||||||||
Db   73049 CCTCTGCATTATAGCCTGGTTCTTCATCATTGGTCTGGTGGGAAACTTGTTGGCCTTAGTTG 72990

Qy     161 TTACCTGTCAAAAGAGCAGGAAGAGATCAACTGCACAGGAGCATCTACCTGGTGCACCTGGCTG 220
           |    || |||| ||  |  ||  ||  |||||  ||||||||||||||||||||||
Db   72989 TCATTGTTCAAAATAGAAAAAAAAATCAACTCAACCACTCTCATTCAATGAACTTGGTGA 72930

Qy     221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTCGTTGTTATGTGCTGGGCT 280
           ||| |  |||| ||| ||||  || ||| ||| ||||| ||  |  || ||| || |  |
Db   72929 TTTCCGACATCCTGTTTACCTACCGCGGATAGTCTTACTTATGCCCTGGGTT 72870

Qy     281 CCAGCTGGCCTTTCGGCAAGGGGCTCGAGGCTGACGGCGTTTGTGCTCTACACCGACA 340
           || ||||| |   |||  ||||| || ||||||| |  | ||||| | || ||| |
Db   72869 TTGACTGGAGGATTGGTGATGCCCTGTGCCGGATAACTGCTCTGTGTTCTACATCAACA 72810

Qy     341 CCTACGGGGGGTCTACCTCATGGCCTGTGTGAGGTGGACCATTACCCAGCTGTGGTCT 400
           || || ||  |   ||||  |||  || || || |||||| ||  ||| || ||||||
Db   72809 CGTATGCAGGTGTGAACTTCATGAACTTGTTGAGCATAGACCGATTCTTTGCCGTGGTGC 72750

Qy     401 GTGCCCACTGGGCCCGCCGCCTCCGCGACACTGCGGCTGGCTGGTCGCGTGGCCA 460
           || ||||||||| ||||| ||| ||  |  ||| ||| |||| | || |  ||| |||
Db   72749 ACCCTCGCGCTACAACAAGATTAAAAGAATTGAATATGCAAAGGGCATTGTGTATTCTG 72690

Qy     461 TCTGACCTTGGTGCTGCTGCCGCAGCGATGGCCCGTTGCTCCTTGATGCCCATGACCAAGCCGC 520
           |||| ||||||||| ||||||||| ||||||  | |  ||||| ||||| || | || ||
Db   72689 TCTGGATTTTGGTGTTTGCTCAAACCATGCCACTGCCTCTCAAACCCATGTCTAAGCAGG 72630

Qy     521 TGGTGGCAAGCTGGCTGCTGCATGGAGCTACAGCAGGTACAGCAGCATGGAGTCGTCCTCGGGCTGCCCC 580
           |||||| ||||||||||||| || || || ||||| || ||  |||| ||||| || ||
Db   72629 AGGCAGACAAGACTACTTGCATGGAATATCCAAATTTGAAGGGACGACATCTCTCCCGT 72570

Qy     581 TCATGTCCTGGTGGCCTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCGTCCT 640
           ||||| || ||  | || ||| || || ||||| |||| | |  ||||| |  |||||
Db   72569 GGATTCTGCTCGGAGCCTGCCTAGCTGCACTCGGCAGCGTCGGGAGGACCCAGTGACCAGCA 72510

Qy     641 GCTATATGAAGATCACCTGGAAGCTGTGCGAGCACAGCTCGGAGGACACCAGTGACCAGCA 700
           || || |||| || ||| |||||| ||| ||| ||||||||  ||  || |  |  ||
Db   72509 GCTACTCTCAGATCTGCTGCAAACTCTTCAGGACTCCCAAGCAGCAGACCCACTGACCGAGA 72450

Qy     701 GGAAAGGACGGCCACTGCGGAGGCTGCGAGGCTGGCTGGTGCCGTGGTGGCCGTGGTGTGG 760
           ||||||||| |||  || ||  | ||| |||| |||| || || || || |||| ||  |
Db   72449 AATCTGGTGTGAACAAAAAGGCTCTCAACACGATCATCCTCAAGCAGTTCATGGGCGAGAGGATGTCGTCT 72390

Qy     761 TCTGCTTCAGGCCTACCACCTCAACATCAAGCAGTTCATGGGCGAGAGGATGTCGTCACC 820
           |||||||||  | ||||||||||  ||| ||| |  |||| ||  || ||| | || |
Db   72389 TGTGCTTCACGCGCCATTACCACGTGGCCATTATGCAGCACATGGTGAGACTCTCTAGCCCC 72330

Qy     821 TGCCATC-------CTGTGCCGAGCGGAGGGCTTCTTCACTGTCCCTTCAGGCCACCGTGG 874
           |||| |       || | ||||||  |||||| || || || | ||||| || |  ||
Db   72329 CCGGAGCCGTGGGGTGAGACTCTTCCCGACTCTCTGCACTTCATCCACTTCACACAGTGT 72270

Qy     875 CCCTCATGAACATGAACTTGTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
           ||| || ||||| ||| || ||| | ||| |  |||  ||||  || | |||  || ||
Db   72269 GCCTGATGAACTTCAACTGCTGCATGGACCCCTTTCATATATACTTCTTCTTTGCATGTAAAGGGT 72210
```

Mon Feb 23 11:38:27 2004   us-10-041-615-33.rge

```
Qy      935  ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCC 987
             ||||||  |||| |||||  ||| ||  |  ||| |||||||||||||||||
Db    72209  ACAAGAGAAAGGTCATGAAGATGCTCAAACGTCAAGTGAGCGTGTCAATCTCC 72157
```

RESULT 13
AC094739/c
LOCUS   AC094739   244625 bp   DNA   linear   HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-5C8, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC094739
VERSION   AC094739.5 GI:30466566
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 244625)
AUTHORS   Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen.H., Alsbrooks,S., Amin.A., Anguiano,D.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,M., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto.M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey.Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mallory,K., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,M., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarrpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Perez,J., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,M., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,J., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,J., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 244625)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 244625)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 9, 2003 this sequence version replaced gi:22771707.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contig within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GBGQ
               Center clone name: CH230-5C8
            -------------- Summary Statistics
               Assembly program: Atlas;
               Consensus quality: 227319 bases at least Q40
               Consensus quality: 230441 bases at least Q30
               Consensus quality: 232587 bases at least Q20
               Estimated insert size: 238330; sum-of-contigs estimation
               Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/genbank draft data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. Gaps between the contigs
            *   are represented as runs of N. The order of the pieces
            *   is believed to be correct as given, however the sizes
            *   of the gaps between them are based on estimates that have
            *   provided by the submittor.
            *   This sequence will be replaced
            *   by the finished sequence as soon as it is available and
            *   the accession number will be preserved.
            *   1 244625; contig of 244625 bp in length.
FEATURES   Location/Qualifiers
     source   1..244625
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-5C8"
     misc_feature   1..1336
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature   1873..5771
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH335182"
     misc_feature   45659..47268
                     /note="wgs_contig"
     misc_feature   complement(242705..243403)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH334721"
ORIGIN
```

Query Match   18.9%;   Score 208.6; DB 2; Length 244625;
Best Local Similarity 53.0%; Pred. No. 3.4e-32;
Matches 473; Conservative 0; Mismatches 414; Indels 6; Gaps 1;

```
Qy    101 CTCTGTTCTACACAGCCCTCCTGCTCCTGCGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
Db 100918 CTCTGCATTATAGCCTGGTCTTCATCATTGGTCTGGTGGGAAACTTGTTGGCCTTAGTTG 100859

Qy    161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCACTGTGCACCTGGCTG 220
Db 100858 TCATTGTTCAAAATAGAAAAAAATCAACTCACCACCCTCTATTCAATGAACTTGTGA 100799

Qy    221 TGTCTGACCTGCTGTTCACCGTGCCTTACCGGGAAGGGTGGTGTGTTATGTGCTGGGCT 280
Db 100798 TTTCCGACATCCTCGTTTACCACAGCTTTACCACGCGGATAGTCCTTACCACGCCCTGGGTT 100739

Qy    281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGGCGGTTGTGCTCAGCCGTTGTGCTTACCACCGACA 340
Db 100738 TTGACTGGAGGATTGGTGATGCCGGTCGACGGGCGGTTGTGCTCGTGTTTACATCAACA 100679

Qy    341 CCTACGGGGGGGTCTGCATGGAGTACAGCAGCAGCAGCCATTACCCAGCTGTGGTCT 400
Db 100678 CGTATGCAGGTGTGAACTTCATGCATAGACCGATTCTTGCCGTGTGTGC 100619

Qy    401 GTGCCCACTGGGGCCCGCGCTGCCCGACAGGCTGGTCTCGCGTGCCA 460
Db 100618 ACCCTCTACGCTACACAAGATAAAAAGGCATTGTGTATTTG 100559

Qy    461 TCTGGACCTTGGTGCTGCGAGACAGTTCAGGATGCCCTGGCAGCCGGC 520
Db 100558 TCTGGATTGTTGGTTGGTCAAACACTGCCTCCTCAAACCCATGTCTAAGCAGG 100499

Qy    521 TGGTGGGCAAGCTGGCTGGCCAGCATGGAGTCAGTCGGGGGTGTTCGTCT 580
Db 100498 AGGCAGACAAGACTACTTGCATGGAATATCAAATTGAAGGAACAGCATCTCCCGT 100439

Qy    581 TCATGGTCCTGGTTGGTGGGCCAGTGGACTGCTGCCTCTTGCCATCATCCTGT 640
Db 100438 GGATTCTTCGGATTCTGCCTAGGCTACGTGGCTGCTTGCCATCATTCCTGT 100379

Qy    641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGTCGGGAGGACCAGTGACCAGCA 700
Db 100378 GCTACTCTCAGATCGTGCAAACTCTTCAGGACGTGCCAAGCAGAACCCACTGACCGAGA 100319

Qy    701 GGAAAGGACGACCCACTGGGCGAGGCGTCGCTGCTGCCTGCGTGTGGGCCGTGTGG 760
Db 100318 AATCTGGTGTGAACAAAAAGGTCTCAAACGATCATCCTCATCGGCGTGTTCGTCT 100259

Qy    761 TCTGCTTCAGCCCCTACCACCTTCAGTTCATGGCGAGGAGGATGCTCCACC 820
Db 100258 TGTGCTTCACGCCCTACCACGTGGCCATTATGGCACACTGGTGAAGACTCTCTACGCCC 100199

Qy    821 TGCCATC------CTGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db 100198 CCGGAGCCCCTCGGCTGTGCGGGGTGAGACACTCCTTCCAGATCTCTTCACTTCACAGTGT 100139

Qy    875 CCCTCATGAACATGAACTGTGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
Db 100138 GCCTGATGAACTTCAATCGCTGCATGAGATGCTGAAACTGTAAAGGGT 100079

Qy    935 ACAGGAAAATGGCTCCTGGGCATTTAAAGCTCAAAAGGTCTTCCTCCCTCC 987
Db 100078 ACAAGAGAAAGGTCATGAAGAAGGTCAAACGTCAAGTGAGCGTGTCAATCTCC 100026
```

Query Match   16.7%;   Score 184.8; DB 6; Length 1086;
Best Local Similarity 50.6%; Pred. No. 2.1e-27;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

```
Qy 101 CTCTGTTCTACACAGCCCTCCTGCTCCTGCGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
Db 104 CTCTGCATTACAGCCCTCGGTCTTCATCATTGGTCTGGTGGGAAACTTGTTGGCCTTAGTTG 163

Qy 161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCACTGTGCACCTGGCTG 220
Db 164 TCATTGTTCAAAATAGAAAAAAATCAACTCACCACCCTCTATTCAATGAACTTGTGA 223

Qy 221 TGTCTGACCTGCTGTTCACCGTGCCTTACCGGGAAGGGTGGTGTGTTATGTGCTGGGCT 280
Db 224 TTTCTGATATATACTTTTTACCACGCGGATAGTCCTTACCACGCAATGGGCT 283

Qy 281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGGCGTTTGTGCTCAGCCGTTGTGCTCTTACACCGACA 340
Db 284 TTGACTGAGAATCGGAGATGCCTTGTGTAGGATAACTGCGCGTAGTGTTTACATCAACA 343

Qy 341 CCTACGGGGGGTCTACCTCATGGAGTACAGCAGCAGCATGAGTCGACAGCTGGTCT 400
Db 344 CATATGCAGGTGTGAACTTCATGACCTTTATGACCTGAGTATTGACCGCTTCATTGCTGTGTGC 403

Qy 401 GTGCCCACTGGGGCCCGCGCTGCCCGACAGGCTGGTCTCGCGTGCCA 460
Db 404 ACCCTCTACGCTACACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG 463

Qy 461 TCTGGACCTTGGTGCTGCGAGACAGTTCATGATGCCCTGGCTCTCGTGATGACCAAGCCGC 520
Db 464 TCTGGATTCTAGTATTGCTCAGACACTCCACTCCTCATCAACCTATGTCAAAGCAGG 523

Qy 521 TGGTGGGCAAGCTGGCTGGCGACAGCAGCATGGAGTCAGTCGCTGCGCCCC 580
Db 524 AGGCTGAAAGGATTACATGCATGGATGTATCCAAATTTGAGAGAAACTAAATCTCTCCCT 583

Qy 581 TCATGGTCCTGGTTGGTGGGCCAGTGACTGCTGCCTCTTGCCATCATCCTGT 640
Db 584 GGATTCTGCTTGGGCATGTTCATAGGATATGTACTTCCACTTATAATCATTCATCATCT 643

Qy 641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGTCGGGAGGACCAGTGACCAGCA 700
Db 644 GCTATTCTCAGATCGTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 703

Qy 701 GGAAAGGACGCCACTGGGCGAGGCGTTGCTGCCTGCGTGTGGGCCGTGTGG 760
Db 704 AATCTGGTGAACAAAAAGGCTCTCAACAATTATTCTTATTATTGTTGTTGTTC 763

Qy 761 TCTGCTTCAGCCCCTACCACCTTCATGGCGGAGGATGCTCCACC 820
Db 764 TCTGTTTCACACTTACCATGTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 823

Qy 821 TGCCATC------TGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db 824 CTAATTTCCTGAATGAGCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 883

Qy 875 CCCTCATGAACATGAACTGTGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
Db 884 GCCTGATGAACTTCAATTGCTGCATGAGACCCTTTTATCTACTTCTTTGCATGTAAAGGGT 943
```

RESULT 14
AR308611

```
Qy    935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCT 994
Db    944 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1003
Qy    995 CCTCCTCCAGGAAAAGCTTCTTCAGAAACACCA 1032
Db   1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041
```

RESULT 15
AX704977
LOCUS   AX704977   1086 bp   DNA   linear   PAT 04-APR-2003
DEFINITION Sequence 1 from Patent EP1286165.
ACCESSION   AX704977
VERSION   AX704977.1 GI:29561638
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Fidock,M.D. and Robas,N.M.
  TITLE   Neuropeptide receptor and uses thereof
  JOURNAL   Patent: EP 1286165-A 1 26-FEB-2003;
            Pfizer Limited (GB); PFIZER INC. (US)
FEATURES   Location/Qualifiers
     source   1..1086
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   16.7%; Score 184.8; DB 6; Length 1086;
Best Local Similarity   50.6%; Pred. No. 2.1e-27;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

```
Qy   101 CTCTGTTCTACACAGCCTCCTCCTGGTCTTCAGTGCCCTCGGGGAAACATCCTTGCCCTTTGCC 160
Db   104 CTCTGCATTACAGCCTCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 163
Qy   161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
Db   164 TCATTGTTCAAAACAGGAAAAAAAATCAACTCTACCACCCTCATTCAACAAATTGGTGA 223
Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCCGGAAGGTGGTGTTATGTGCTGGGCT 280
Db   224 TTTCTGATATACTTTTTACCACGGCTTGCCTCACAGCGAATAGCCTACTATGCAATGGGCT 283
Qy   281 CCAGCTGGCCTTTGGCAAGGGGCCTCGCAGGCGTGACGGCGTTTGTGCTCTACACCGACA 340
Db   284 TTGACTGAGAATCGGAGATGCCTTGTGTGAGGATAAACTGCGCGCTAGTGTTTTACATCAACA 343
Qy   341 CCTACGGGGGGGTCTACCCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
Db   344 CATATGCAGGTGTGAACTTTATGACCTGCCCTGAGTATTGACCGCTTCATTGCTGCTGGTGGTGC 403
Qy   401 GTGCCCACTGGGGCGCCGCCGGCCCCGGCACCGCTGGCGCCCAGCGTGGTCTCGCGTGCGCCA 460
Db   404 ACCCTCTACGTACAACAACAAGATTGAACATGCAAAAGGCGTGTGCATATTTG 463
Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTTGATGCCCATGACCAAGCCGC 520
Db   464 TCTGGAATCTAGTATTTGCTCAGACACTTCCACTCCATCAACCCTATGTCAAAGCAGG 523
Qy   521 TGGTGGGCAAGCAGCTGCCTGACATGGAGTACGAGCAGCATGGAGTCCTGGGCTGCCCC 580
Db   524 AGGCTGAAAGATTACATGCATGGAGTAATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 583
Qy   581 TCATGGTCCTCGGTGGCCTTTGCCATTGGCCTTCTGTGGGCCAGTGGGATCATCCTGTCCT 640
Db   584 GGATTCCGCTTGGGGCATGTTTCATCAGGATATGTGACTTCCACTTATATAATCATTCTCATCT 643
Qy   641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACTGGCGGAGGACCTCGGGAGGACCCCAGTGACCAGCA 700
```

```
Db   644 GCTATTCTCAGATCTGCTGCAACTCTTGCGAACTCTTCAGAACTGCCAAACAAACCCACTCACTGAGA 703
Qy   701 GGAAAGGACGCCACTGGCGGAGGCTGCCTGCTTACGCTGCTCATGCTGGTCGGCCGTGGGTGG 760
Db   704 AATCTGGTGTAAACAAAAAGCTCTCAACACAATTATTCTTATTATTGTTGTGTTTGTTc 763
Qy   761 TCTGCTTCAGCCCTCACCCTTCAAATCAAGCAGTTCATGGCGAGAGGATGCTCCACC 820
Db   764 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 823
Qy   821 TGTCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db   824 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 883
Qy   875 CCCTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
Db   884 GCCTGATGAACTTCAATTGCTGCATGGACCCTTTATCTACTTCTTTGCATGTAAAGGGT 943
Qy   935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCT 994
Db   944 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1003
Qy   995 CCTCCTCCAGGAAAAGCTTCTTCAGAAACACCA 1032
Db  1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041
```

Search completed: February 21, 2004, 20:39:39
Job time : 4349 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         February 21, 2004, 17:32:38 ; Search time 484 Seconds
                                    (without alignments)
                              9690.105 Million cell updates/sec

Title:          US-10-041-615-33
Perfect score:  1104
Sequence:       1 atgataaactaggcccta.........gtgtcatttgctttatgact 1104

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_29Jan04:*
                 1: geneseqn1980s:*
                 2: geneseqn1990s:*
                 3: geneseqn2000s:*
                 4: geneseqn2001as:*
                 5: geneseqn2001bs:*
                 6: geneseqn2002s:*
                 7: geneseqn2003as:*
                 8: geneseqn2003bs:*
                 9: geneseqn2003cs:*
                10: geneseqn2004s:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1104	100.0	1104	6	ABS64723	Abs64723 cDNA enco
2	779	70.6	4319	9	ADC86840	Human GPC
3	779	70.6	9409	4	AAK76477	Human imm
4	672.8	60.9	1526	6	ABK94922	Human nov
5	655	59.3	1771	5	AAS78869	Human DNA
c 6	594.4	53.8	1001	5	AAS98131	DNA encod
7	397.2	36.0	536	4	AAS42840	Human G P
c 8	378.4	34.3	556	6	AAS30791	Human cDN
c 9	328.4	29.7	330	6	AAS98111	Human DNA
10	184.8	16.7	1086	3	AAA30635	Human G p
11	184.8	16.7	1086	9	ACC48996	Human G-p
12	184.8	16.7	1086	9	ADC22596	Human G p
13	184.8	16.7	1643	2	AAQ64126	Epstein B
14	184.8	16.7	1643	2	AAV25491	cDNA for
15	184.8	16.7	1643	7	ACA56681	Human sig
16	184.8	16.7	1643	9	ADE84890	Farnesyl
17	184.8	16.7	2477	9	ADE25635	Human cDN
18	183.2	16.6	1086	3	AAA30730	DNA encod
19	183.2	16.6	1086	9	ADC22724	Human G P
20	183.2	16.6	1638	7	AB242660	Human EBV
21	172.2	15.6	1556	3	AAA52629	Eosinophi
c 22	148	13.4	594	6	AAS98061	Human DNA
23	123.4	11.2	548	8	ACH23286	Human adu

24	119.4	10.8	1020	6	ABQ79300	Human GPC
25	119.4	10.8	1076	6	AAD29667	Human G-p
26	118.6	10.7	2025	6	AAD04980	Human pur
27	118.6	10.7	2025	6	ABZ35611	Human gen
28	118.6	10.7	2025	7	ACA56884	Human sig
29	118.6	10.7	2118	6	ABZ42713	Human pur
30	117	10.6	2118	6	ABZ35320	Human gen
31	116	10.5	1163	4	AAD04981	Human pur
32	115	10.4	1017	9	ADC25999	Human pur
33	115	10.4	1017	9	AAD61648	Human GPC
34	112	10.1	1020	3	AAA30601	Human G p
35	112	10.1	1020	3	AAA30718	DNA encod
36	112	10.1	1020	6	ABK52848	Human cys
37	112	10.1	1020	9	ADC22550	Human G p
38	112	10.1	1020	9	ADC22700	Human G p
39	112	10.1	1900	7	ACA56886	Human sig
40	112	10.1	1901	2	AAQ66178	Seven tra
41	112	10.1	1901	2	AAV18356	Human R12
42	112	10.1	1901	3	AAA91725	Human 7TM
43	112	10.1	1901	6	ABK54254	Human 7 t
44	112	10.1	2070	6	ACC72705	Human can
45	112	10.1	2070	7	ABZ42757	Human G p

```
                              ALIGNMENTS

RESULT 1
ABS64723
ID   ABS64723 standard; cDNA; 1104 BP.
XX
AC   ABS64723;
XX
DT   15-NOV-2002  (first entry)
XX
DE   cDNA encoding human GPCR16 protein.
XX
KW   G protein coupled receptor related protein; ss; gene; human; GPCR;
KW   cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
KW   von Hippel-Lindau syndrome; Alzheimer's disease; tuberous sclerosis;
KW   hypercalcaemia; Parkinson's disease; Huntington's disease;
KW   cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW   ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW   pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW   inflammatory bowel disease; infertility; hypertension; scleroderma;
KW   arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW   infection; graft-versus-host disease.
XX
OS   Homo sapiens.
XX
PN   WO200264793-A2.
XX
PD   22-AUG-2002.
XX
PF   03-JAN-2002; 2002WO-US000056.
XX
PR   03-JAN-2001; 2001US-0259552P.
PR   09-JAN-2001; 2001US-0260544P.
PR   20-MAR-2001; 2001US-0277405P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Casman SJ, Edinger SR, Ellerman K, Smithson G, Kekuda R;
PI   Muralidhara P;
XX
DR   WPI; 2002-643487/69.
DR   P-PSDB; ABG33286.
XX
PT   New isolated G protein coupled receptor polypeptides and polynucleotides,
PT   useful in gene therapy, particularly for treating or preventing
PT   cardiomyopathy, atherosclerosis, diabetes, multiple sclerosis, Crohn's
PT   disease or cancer.
XX
```

```
PS   Claim 9; Page 97; 252pp; English.
XX
CC   This invention relates to a new isolated G-protein coupled receptor
CC   (GPCRX) polypeptide sequence and the cDNA encoding it. The GPCRX
CC   polypeptide, GPCRX nucleic acid and an antibody specific to the protein
CC   are useful for treating, preventing or alleviating a GPCRX-associated
CC   disorder or a pathological state in a subject, particularly a human. In
CC   particular, the disorder is cardiomyopathy, atherosclerosis, diabetes, or
CC   a disorder related to cell signal processing and metabolic pathway
CC   modulation. The GPCRX polypeptide and nucleic acid are also useful for
CC   diagnosing the presence of or predisposition to a disease associated with
CC   altered levels of GPCRX, particularly cancer. The GPCRX nucleic acid and
CC   polypeptide are especially useful in the manufacture of a medicament for
CC   therapeutic or prophylactic applications for disorders associated with
CC   aberrant GPCRX expression or activity, e.g. Von Hippel-Lindau syndrome,
CC   Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia,
CC   Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy,
CC   Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,
CC   leukodystrophies, addiction, anxiety, depression, pain, obesity, Crohn's
CC   disease, osteoporosis, inflammatory bowel disease, infertility,
CC   hypertension, scleroderma, haemophilia, asthma, arthritis, human
CC   immunodeficiency virus, autoimmune disease; HIV, viral, fungal, bacterial
CC   or protozoal infections, or graft-versus-host disease. The DNA encoding
CC   the protein is useful in gene therapy for treating the above conditions.
CC   The polypeptides can be used as immunogens to produce antibodies and as
CC   vaccines. The nucleic acids are further used as hybridisation probes, in
CC   chromosome mapping, tissue typing, preventive medicine, and
CC   pharmacogenomics. These are also useful in developing powerful assay
CC   system for functional analysis of various human disorders, as well as in
CC   diagnostic applications. The present sequence represents the cDNA
CC   encoding a human G protein coupled receptor related protein (GPCR) of the
CC   invention
XX
SQ   Sequence 1104 BP; 217 A; 328 C; 290 G; 269 T; 0 U; 0 Other;

Query Match             100.0%;  Score 1104;  DB 6;  Length 1104;
Best Local Similarity   100.0%;  Pred. No. 6.8e-254;
Matches 1104;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATGATAAAACTAGGCCCTTACTTCACACACCACCACAAAAATAAAAAACTAAGAATCTAAGAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGATAAAACTAGGCCCTTACTTCACACACCACCACAAAAATAAAAAACTAAGAATCTAAGAT 60

Qy     61 TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGGCGTCCTCCTGTTCTCACACAGCCCTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGGCGTCCTCCTGTTCTCACACAGCCCTC 120

Qy    121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTTACCTGTCAAAAGAGCAGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTTACCTGTCAAAAGAGCAGG 180

Qy    181 AAGATCAACTGACACAGGCATCACCTGGTGCACCTGGTCTCGACCTGTGTTCACC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGATCAACTGACACAGGCATCACCTGGTGCACCTGGTCTCGACCTGTGTTCACC 240

Qy    241 GTGGCCTTACCGGGAAGGGTGGTGTTATGTGGGCTCCAGCTGGCCTTTCGGCAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTGGCCTTACCGGGAAGGGTGGTGTTATGTGGGCTCCAGCTGGCCTTTCGGCAAG 300

Qy    301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCACGGGGGGGGTCTACCTC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCACGGGGGGGGTCTACCTC 360

Qy    361 ATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGCCGCGGC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGCCGCGGC 420

Qy    421 CTCCGCACGGCTGGCCGCCGCCAGGCTGGTCTCGCGGCCATCTGGCCGCCTGGACCTGGTCTGTGTGCTG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCCGCACGGCTGGCCGCCGCCAGGCTGGTCTCGCGGCCATCTGGCCGCCTGGACCTGGTCTGTGTGCTG 480

Qy    481 CAGACAGCATGCCCTGCTTCTTGATGCCCATGACCAAGCTGGTGGGCAAGCTGCCTGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGACAGCATGCCCTGCTTCTTGATGCCCATGACCAAGCTGGTGGGCAAGCTGCCTGC 540

Qy    541 ATGGAGTACAGGACGAGCATGGAGTCAGTCGTCCTCGGGCTGCCTGGCCTTT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGGAGTACAGGACGAGCATGGAGTCAGTCGTCCTCGGGCTGCCTGGCCTTT 600

Qy    601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG 660

Qy    661 AAGCTGTGCAGCACAGTCGGGAGGACCAGTGACCAGCAGGAAAGGACGCCACTGGCGA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AAGCTGTGCAGCACAGTCGGGAGGACCAGTGACCAGCAGGAAAGGACGCCACTGGCGA 720

Qy    721 GGCTGCCTGCTTACGCTGCGTATGCTGGTGGCCGTGGTGGTCTGCTTCAGCCCTACCAC 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GGCTGCCTGCTTACGCTGCGTATGCTGGTGGCCGTGGTGGTCTGCTTCAGCCCTACCAC 780

Qy    781 CTCAACATCAAGCAGTTCATGGCGAGGAGGATGTCCACCTGGCCATCCTGTGCCGAGCGG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTCAACATCAAGCAGTTCATGGCGAGGAGGATGTCCACCTGGCCATCCTGTGCCGAGCGG 840

Qy    841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCGGTGGCCCTCATGAACATGAACTGTGGCATT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCGGTGGCCCTCATGAACATGAACTGTGGCATT 900

Qy    901 ACCCCAATCATTACTTCTTTGCATCCACCCATTACAGGGAAAATGGCTCCTCGGGCATTTTA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACCCCAATCATTACTTCTTTGCATCCACCCATTACAGGGAAAATGGCTCCTCGGGCATTTTA 960

Qy    961 AAGCTCAAAGGGTCTTCCTCCTCTCCTCCTCCTCCTCATGTTCTTAGCGGAGCACGTGGTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGCTCAAAGGGTCTTCCTCCTCTCCTCCTCCTCCTCATGTTCTTAGCGGAGCACGTGGTC 1020

Qy   1021 TCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC 1080

Qy   1081 TAACGTGTCATTTGCTTTATGACT 1104
          ||||||||||||||||||||||||
Db   1081 TAACGTGTCATTTGCTTTATGACT 1104

RESULT 2
ADC86840
ID   ADC86840 standard; DNA; 4319 BP.
XX
AC   ADC86840;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1293.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86841.
```

```
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1293; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 4319 BP; 1058 A; 1049 C; 1170 G; 994 T; 0 U; 48 Other;

Query Match             70.6%;  Score 779;  DB 9;  Length 4319;
Best Local Similarity   83.6%;  Pred. No. 5.6e-176;
Matches 1006;  Conservative   0;  Mismatches   0;  Indels  197;  Gaps   3;

Qy     97 CGCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2940 CGGTCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 2999

Qy    157 TGCCTTACCTGTCAAAGAGACAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3000 TGCCTTACCTGTCAAAGAGACAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 3059

Qy    217 GCTGTGTCTGACCTGCTGTTCACCGGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTG 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3060 GCTGTGTCTGACCTGCTGTTCACCGGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTG 3119

Qy    277 GGCTCCAGCTGGCCTTCGGCAAGGGCTCTGCAGGGCTGACGGCGTTGTGCTCACACC 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3120 GGCTCCAGCTGGCCTTCGGCAAGGGCTCTGCAGGGCTGACGGCGTTGTGCTCACACC 3179

Qy    337 GACACCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 396
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3180 GACACCTAC-GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 3238

Qy    397 GTCTGTGCCCACTGGGACCTGCGCCAGGCTGGTCTGCGTG 456
          ||| |||||||||||||||||||||||||||||||||||
Db   3239 GTC-GTGTGCCCACTGGGACCTGCGCCAGGCTGGTCTGCGTG 3298

Qy    457 GCCATCTGGACCTGCGCAGAGATGATGCCCTTGCCTCTTGATGACCAAG 516
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299 GCCATCTGGACCTGCGCAGAGATGATGCCCTTGCCTCTTGATGACCAAG 3358

Qy    517 CCGGCTGGTGGGCAAGCTGCCTGCATGGAGCAGCAGCATGGAGTCAGTCCTCCGGGCTG 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359 CCGGCTGGTGGGCAAGCTGCCTGCATGGAGCAGCAGCATGGAGTCAGTCCTCCGGGCTG 3418

Qy    577 CCCTCATGTGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGTGGGCCAGTGGGGATCATCCTG 636
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419 CCCTCATGTGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGTGGGCCAGTGGGGATCATCCTG 3478

Qy    637 TCCTGCTATATGAAGATCACCTGAAGCTGTGCGCACAGCT------------- 678
          ||||||||||||||||||||||||||||||||||||||
Db   3479 TCCTGCTATATGAAGATCACCTGAAGCTGTGCGCAGCTGTGCGAGCACTGGGAGACCCAGTGACCA 3538

Qy    679 ---------------------- 678
Db   3539 GCGGGAAAGGACACCACCGGCGGGGCGCAGCCCAGGAGGACCCAGTGACCAACAAGGAGGAAAGGA 3598

Qy    679 ---------------------- 678
Db   3599 CGCCACCGGCGGGGCGCAGCCCAGGAGGACCCAGTGACCAGCGGGAAAGGACACCACCGGCA 3658

Qy    679 ------------------CGG 681
                            |||
Db   3659 GGACAGCCCGAGGAGGACCCAGTGACCAGCAGGAAAGGACGGCCACCGGCGGGGCGGGGACAGCAGCCCGG 3718

Qy    682 GAGGACCCAGTGACCAGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTACCGCTGCTG 741
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3719 GAGGACCCAGTGACCAGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTACCGCTGCTG 3778

Qy    742 ATGCTGGTGGCCGTGGTGGTCGTCGCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATG 801
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3779 ATGCTGGTGGCCGTGGTGGTCGTCGCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATG 3838

Qy    802 GCGAGAGGGATGCTCCACCTGCCATCCTGTCGCCGAGCGGAGGCGGCTTCTTACTGTCCCTT 861
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3839 GCGAGAGGGATGCTCCACCTGCCATCCTGTCGCCGAGCGGAGGCGGCTTCTTACTGTCCCTT 3898

Qy    862 CAGGCCACCGTGGCCCTCATGGAACATGAACATGTGGCATTACCCCAATCATTTACTTCTTT 921
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   3899 CAGGCCACCGTGGCCCTCATGGAACATGAACATGTGGCATTACCCC-ATCATTTACTTCTTT 3957

Qy    922 GCATCCACCCATTACAGGAGAAATGGCTCCTGGGCATTTTAAAGCTCAAAGGGTCTTCCTCC 981
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3958 GCATCCACCCATTACAGGAGAAATGGCTCCTGGGCATTTTAAAGCTCAAAGGGTCTTCCTCC 4017

Qy    982 TCCTCCTCCTCCTCCTCCTCCTCCACCCCAGAAAAAGCTTCTTCAGAAAACACCAAGTATCACC 1041
          |||||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Db   4018 TCCTCCTCCTCCTCCTCCTCCTCCCCAGGAAAAGCTTCTTCAGAAAACACCAAGTATCACC 4077

Qy   1042 CAGGCCAGAGTTCTTATGTTCTTAGCGGAGCCACGTGGTCTAACGTGTCATTTGCTTTATG 1101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4078 CAGGCCAGAGTTCTTATGTTCTTAGCGGAGCCACGTGGTCTAACGTGTCATTTGCTTTATG 4137

Qy   1102 ACT 1104
          |||
Db   4138 ACT 4140


RESULT 3
AAK76477
ID   AAK76477 standard; DNA; 9409 BP.
XX
AC   AAK76477;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31289.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
```

```
PR   14-AUG-2000;   2000US-0224518P.
PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232080P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.

PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI;  2001-48326/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 31289; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54950 and AAM82169
```

```
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 9409 BP; 2123 A; 2163 C; 2259 G; 2864 T; 0 U; 0 Other;

     Query Match    70.6%;  Score 779;  DB 4;  Length 9409;
     Best Local Similarity  83.6%;  Pred. No. 7.3e-176;
     Matches 1006;  Conservative  0;  Mismatches  0;  Indels 197;  Gaps  3;

Qy     97  CGCTCCTCGTTCTACACAGCCCTCGGTCTTCAGTGCCCCTGGGGAAACATCCTTGCCCTT  156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373  CGCTCCTCGTTCTACACAGCCCTCGGTCTTCAGTGCCCCTGGGGAAACATCCTTGCCCTT  432

Qy    157  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTCACCTGGTGCACCTG  216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTCACCTGGTGCACCTG  492

Qy    217  GCTGTGTCTGACCTGCTGTTCACCGGTGGCCTTACCGGGGAAGGTGGTGTGTTATGTGCTG  276
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493  GCTGTGTCTGACCTGCTGTTCACCGGTGGCCTTACCGGGGAAGGTGGTGTGTTATGTGCTG  552

Qy    277  GGCTCCAGCTGCCCTTTCGGCAAGGGGGCTCTGCAAGGCTGACGGCCGTTTGTGCTCTACACC  336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553  GGCTCCAGCTGCCCTTTCGGCAAGGGGGCTCTGCAAGGCTGACGGCCGTTTGTGCTCTACACC  612

Qy    337  GACACCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCTGTGACCATTACCCAGCTGTG  396
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    613  GACACCTAC-GGGGGTCTACCTCATGGCCTGTGTGAGCTGTGACCATTACCCAGCTGTG  671

Qy    397  GTCTGTGCCCACTGGGGGCCCCGGCCTCCGCACGGCTGGCCGCGGCTGGTCTCGCGTG  456
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672  GTCTGTGCCCACTGGGGGCCCCGGCCTCCGCACGGCTGGCCGCGGCTGGTCTCGCGTG  731

Qy    457  GCCATCTGGACCTTGGTGTCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG  516
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732  GCCATCTGGACCTTGGTGTCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG  791

Qy    517  CCGCTGGTGGGCAAGCTGGCCTGCATGGAGTACAGACAGCATGGAGTCAGTCCTCCGGGCTG  576
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792  CCGCTGGTGGGCAAGCTGGCCTGCATGGAGTCAGCATGGAGTCAGTCCTCCGGGCTG  851

Qy    577  CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGTGGGGATCATCCTG  636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852  CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTTCTGTGTGGGGATCATCCTG  911

Qy    637  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGACAGCT-------------------  678
           |||||||||||||||||||||||||||||||||||||||||
Db    912  TCCTGCTATATGAAGATCACCTGGAAGCTGTGGAGCACAGCTGGCGAGAACCCAGTGACCA  971

Qy    679  ----------------------------------------------------------  678
Db    972  GCGGGGAAAGGACACCACCGGCGGGGGCAGCCCAGGAGGACCCAGTGACCAGACAGGAAGGA  1031

Qy    679  ----------------------------------------------------  678
Db   1032  CGCCACCGGCGGGGGCAGCCCAGGAGGACCCAGTGACCAGCGGCGGAAAGGACACCACCGGCA  1091

Qy    679  -------------------------------------------------CGG  681
                                                            |||
Db   1092  GGACAGCCCCAGGAGGACCCAGTGACCAGCAGGGGAAAGGACGCCACCGGCGGGACAGCCCCGG  1151

Qy    682  GAGGAACCCAGTGACCAGGAGGAAAGGACGACCCACTGGCAGCGGGAAAGGACA  741
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152  GAGGACCCAGTGACCAGCAGGAGGAAAGGACACCACTGGCCTGCTTACGCTGCTG  1211

Qy    742  ATGCTGGTGCCGTGGTGGTCGCTGTGGTTGTCGCTGCTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG  801
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212  ATGCTGGTGGCCGTGGTGGTCGCTGTGGTTGTCGCTGCTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG  1271

Qy    802  GCGAGGGGATGCTTCCACCTGTCCACCTGCCATCCTGTGCCGAGCCGAGGCCTTTCTTCCTTACTGTCCCTT  861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272  GCGAGGGGATGCTCCACCTGTCCACCTGCCCATCCTGTGCCGAGCCGAGGCGGTTTCTTCCTTACTGTCTCCTT  1331

Qy    862  CAGGCCACCGTGGCCCTCATGGAACATGAACATGACTGTGCCATTACCCCAATCATTCATTTACTTCTT  921
```

```
Db   1332  CAGGCCACCGTGGCCCTCATGGAACATGAACATGACTGTGCCATTACCCCAAT-CATTTACTTCTT  1390

Qy    922  GCATCCACCCATTACAGGAGAAATGGCTCCTGGCCATTTAAAGTCAAGGGTCTTCCTCC  981
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391  GCATCCACCCATTACAGGAGAAATGGCTCCTGGCCATTTAAAGTCAAGGGTCTTCCTCC  1450

Qy    982  TCCTCCTCCTCCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCAAGTATCACC  1041
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451  TCCTCCTCCTCCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCAAGTATCACC  1510

Qy   1042  CAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTCTAACGTGTCATTGCTTATG  1101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511  CAGGCCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTCTAACGTGTCATTGCTTTATG  1570

Qy   1102  ACT  1104
           |||
Db   1571  ACT  1573
```

RESULT 4
ABK94922
ID   ABK94922 standard; cDNA; 1526 BP.
XX
AC   ABK94922;
XX
DT   30-AUG-2002 (first entry)
XX
DE   Human novel polynucleotide #33.
XX
KW   Human; gene; ss; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombosis; thrombolysis; coagulation disorder;
KW   fungal infection.
XX
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US047004.
XX
PR   30-NOV-2000; 2000US-00728952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI   Yamazaki V, Ujwal ML, Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   P-PSDB; ABG66698.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders, cancer and promoting wound healing.
XX
PS   Claim 1; Page 403-404; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral

```
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABK94890-ABK94982 represent human
CC   novel polynucleotides of the invention
XX
SQ   Sequence 1526 BP; 282 A; 458 C; 458 G; 328 T; 0 U; 0 Other;

Query Match             60.9%;  Score 672.8;  DB 6;  Length 1526;
Best Local Similarity   81.9%;  Pred. No. 9.7e-151;
Matches  901;  Conservative    0;  Mismatches    2;  Indels  197;  Gaps   3;

Qy         98 GCTCTCTGTTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 157
              ||||||||||| |||||||||||||||||||||||||| ||||||||||||||||||||||
Db        422 GCTCTCTGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCTCTGGGAAACATCCTTGCCCTTT 481

Qy        158 GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGACATCCTTACCTGGTGCACCTGG 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGACATCCTTACCTGGTGCACCTGG 541

Qy        218 CTGTGTCTGACCTGGTCGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTGTTATGTGCTGG 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 CTGTGTCTGACCTGGTCGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTGTTATGTGCTGG 601

Qy        278 GCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGGGCTGACGCGTTTGTGCTCTACACCG 337
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGGGCTGACGCGTTTGTGCTCTACACCG 661

Qy        338 ACACCTACGGGGGGGTCTCACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGG 397
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 ACACCTAC-GGGGGGTCTCACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGG 720

Qy        398 TCTGTGCCCACTGGGGCCCGCCCTCCGCACGGCTGGCCTGCCCAGGCTGGTCTGCGTGG 457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTGTGCCCACTGGGGCCCGCCCTCCGCACGGCTGGCCTGCCCAGGCTGGTCTGCGTGG 780

Qy        458 CCATCTGGACCTTGGTGCTGCGCAGAGCGATGCCCTTGCTCTTGATGACCAAGC 517
              ||||||||||||||||||||| ||||| |||||||||||||||||| ||||||||
Db        781 CCATCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGC 840

Qy        518 CGCTGGTGGGCAAGCTGCCTGCCATGGAGTACAGCAGCATGGAGTCCAGTCCTCGGGCTGC 577
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        841 CGCTGGTGGGCAAGCTGCCTGCCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTGC 900

Qy        578 CCCTCATGGTCCTGGTGGCCTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGT 637
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        901 CCCTCATGGTCCTGGTGGCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGT 960

Qy        638 CCTGCTATATGAAGATCACCTGGAAGCTGTGCGACACAGCT----------- 678
              ||||||||||||||||||||||||||||||||||||| ||||
Db        961 CCTGCTATATGAAGATCACCTGGAAGCTGTGCGACACCAGTGGGAGAACCCAGTGACCAG 1020

Qy        679 ---------------------------------------------- 678
Db       1021 CGGGAAAGGACACCACCGGCGGGGCAGCCCGGGAGGACCCCGGAGGACCAGCAGGAAGGAC 1080

Qy        679 --------------------------------------------- 678
Db       1081 GCCACCGGCGGGGCAGCCCAGTGACCAGCGGGAAAGGACACCACCACCAGCAG 1140

Qy        679 ----------------------------------------CGGG 682
                                                      ||||
Db       1141 GACAGCCCGAGAGGACCCCAGTGACCAGCCCAGGGAAAGGATGCCACCGGCGGGGACAGCCCCGGG 1200

Qy        683 AGGACCAGTGACCAGCAGGAAGGACGCGCCACTGGCGGAGCGGCTGCTGCCTTACGCTGCTGA 742
              ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db       1201 AGGACCAGTGACCAGCAGGAAAGGACGCGCCACTGGCGGAGCGGCTGCTGCCTTACGCTGCTGA 1260

Qy        743 TGCTGGTGGCCGTGGTGGTCGTTCTCAGCCCTACCACCTCAACATCAAGCAGTTCATGG 802
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db       1261 TGCTGGTGGCCGTGGTGGTCGTTCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATGG 1320

Qy        803 CGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTC 862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTC 1380

Qy        863 AGGCCACGTGGCCCTCATGAACATGAACTGGCATTACCCAATCATTACTTCTTTG 922
              ||||||||||||||||||||||||||||||||||||||||| ||||| |||||||||
Db       1381 AGGCCACGTGGCCCTCATGAACATGAACTGGCATTACCCA-ATCATTACTTCTTTG 1439

Qy        923 CATCACCCATTACAGGAAATGGCTCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCT 982
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db       1440 CATCACCCATTACAGGAAATGGCTCTGGGCATTTTAAAGCTCAAAGGGTCTTCCTCCT 1499

Qy        983 CCTCCTCCTCCTCCTCCTCC 1002
              ||||||||||||||||||
Db       1500 CCTCCTCCTCCTCCTCCACC 1519

RESULT 5
AAS78869
ID   AAS78869 standard; cDNA; 1771 BP.
XX
AC   AAS78869;
XX
DT   13-FEB-2002   (first entry)
XX
DE   DNA encoding novel human diagnostic protein #14673.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG14682.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 14673; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences, (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes, (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
```

```
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1771 BP; 313 A; 531 C; 519 G; 407 T; 0 U; 1 Other;

Query Match             59.3%;  Score 655;  DB 5;  Length 1771;
Best Local Similarity   81.4%;  Pred. No. 1.8e-146;
Matches  899;  Conservative   0;  Mismatches   0;  Indels 201;  Gaps   4;

Qy         98 GCTCTCGTTCTACACAGCCCTTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 157
              ||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCTCTCCGTTCTTCACACAGCCCTTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 720

Qy        158 GCCTTACCTGTCA-AAAGAGCAGGAAGATCAACTGACACAGGCATCTACCTGTGCACCTG 216
              |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCCTTACCTGTCAGAGAGAGGAGGAAGATCAACTGACACAGGCATCTACCTGTGCACCTG 780

Qy        217 GCTGTGTCTGACCTGCTTCACCGTGGCCTTACCGGGAAGGTGGTGGTGTTATGTGCTG 276
              |||||||||||||| |||||||||||||||||||||||||||||| |||||||||||||
Db        781 GCTGTGTCTGACCTGGCTTACCGTGGCCTTACCGGGAAGGTGGTGTGTTATGTGCTG 840

Qy        277 GGCTCCAGCTGGCCTTTCGGCAAGGGCTCTGCAGGGGCGTTGTGCTCTACACC 336
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GGCTCCAGCTGGCCTTTCGGCAAGGGCTCTGCAGGGGCGTTGTGCTCTACACC 900

Qy        337 GACACCTACGGGGGGGTTCACTTCATGGCCTGTGTGAGCGTGACCATTACCCAGC-TGT 395
              |||||||||||||||||  |||||||||||||||||||||||||||||||| |||| |||
Db        901 GACACCTACGGGGGGGTCTTCATGGCCTGTGTGAGCGTGACCATTACCCAACTTGT 960

Qy        396 GGTCTGTGCCCACTGGGGCGCCGGGCCTTCCGCACGGCTGGCCCAGGCTGGTCTGCGT 455
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        961 GGTCTGTGCCCACTGGGGCGCCGGGCGTGCCTCGCACGGCTGGCCCAGGCTGGTCTGCGT 1020

Qy        456 GGCCATCTGGACCTTGGTGCTGCTGCACGAGATGCCCTGCTCTTGATGCCCATGACCAA 515
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1021 GGCCATCTGGACCTTGGTGCTGCTGCAGCAGATGCCCTGCTCTTGATGCCCATGACCAA 1080

Qy        516 GCCCTGGTGGGGCAAGCTGGCCTGCCTTGGGTGGCCTTCTGTGGGGCCAGTGGGGATCATCCT 575
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GCCGCTGGTGGGGCAAGCTGGCCTGCCTTGGGTGGCCTTCTGTGGGGCCAGTGGGGATCATCCT 1140

Qy        576 GCCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGTGCGAGCACAGT------ 635
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GCCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGTGCGAGCACAGT 1200

Qy        636 GTCCTGCTATATGAAGATCACCCTGGAAGCTGTGCGAGCACAGCT------ 678
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTCCTGCTATATGAAGATCACCCTGGAAGCTGTGCGAGCACAGCT 1260

Qy        679 ------ 678

Db       1261 AGCGGGAAAGGACCACCACCACCCCGGCGGGCGGGCAGCCCCGGGCAGCCCCGGAGGACCCAGTGACCACCAGAAA 1320

Qy        679 ------ 678

Db       1321 GACGGCCACCCGCGGGACCCCAGGAGGACCCCAGTGACCCCGCGGGAAAGGACCACCACC 1380

Qy        679 ------ 678

Db       1381 AGCAGGACAGCCCGAGAGGACCCAGTGACCAGCGGGAAAGGATGCCACCGGCGGGACAGC 1440

Qy        679 -CGGGAGGAGGAGCCCAGTGACCAGCGGGACCCAGTGACCAGCAGGAAAGACGCGCTGCGAGCTGCTTACGCT 737
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCGGGAGGAGGAGCCCAGTGACCAGCGGGACCCAGTGACCAGCAGGAAAGACGCGCTGCGAGCTGCTTACGCT 1500


Qy        738 GCTGATGCTGGTGGCCGGTGGTGGTCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGT 797
              ||||||||||||||||||||||||||||||||||||||  |||||||||||||||||||||
Db       1501 GCTGATGCTGGTGGCCGGTGGTGGTCTGCTTCAGCCCTTCAGCCCTACCACCTCAACATCAAGCAGT 1560

Qy        798 CATGGCGAGAGGAGGATGCTCCACCTGCACTTCCTGTGCCGAGCGAGGCGGGCTTTCTTACTGTC 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 CATGGCGAGAGGAGGATGCTCCACCTGCACTTCCTGTGCCGAGCGAGGCGGGCTTTCTTACTGTC 1620

Qy        858 CCTTCAGGCCACCGTGGCCCTCATGAACATGAACTGGCATTACCCAATCATTTACTT 917
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1621 CCTTCAGGCCACCGTGGCCCTCATGAACATGAACTGGCATTACCCC-ATCATTTACTT 1679

Qy        918 CTTTGCATCCACCATTACAGGAGAAATGGCTCCTGGCATTTAAAGCTCAAAGGGTCTTC 977
              ||||||||||||| ||||||||||||||||||||| ||||||||||||||||||||||||
Db       1680 CTTTGCATCCACCCATTACAGGAGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTC 1739

Qy        978 CTCCTCCTCCTCCTCCTCCTCC 1002
              |||||||||||||||||||||||
Db       1740 CTCCTCCTCCTCCTCCTCCTCCACC 1764
```

RESULT 6
AAS98131/c
ID   AAS98131 standard; cDNA; 1001 BP.
XX
AC   AAS98131;
XX
DT   12-MAR-2002 (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #88.
XX
KW   Human; G protein-coupled receptor; GPCR; ss; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity; diabetes;
KW   hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US015332.
XX
PR   11-MAY-2000; 2000US-0203217P.
PR   18-MAY-2000; 2000US-0205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin receptor
PT   polypeptides useful for identifying modulators that are useful for
PT   treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT   stroke.
XX
PS   Claim 2; Page 95; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a nucleic
CC   acid molecule that is at least 80% identical to the G protein-coupled
CC   (GPCR) polynucleotides included in the specification. Also included are
CC   probes based on the GPCR sequences (including antisense probes), a host
CC   cell comprising an expression vector comprising the GPCR sequence,

```
CC   antibodies raised against the polypeptides, and methods of identifying
CC   modulators of the polypeptides. The polypeptides are useful for
CC   identifying modulator compounds which function as modulators, activators,
CC   repressors, agonists or antagonists of the novel GPCR polypeptides
CC   including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC   described above can be used to detect the presence of the polypeptides
CC   and nucleic acids and are used to diagnose a variety of diseases or
CC   disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC   amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC   carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC   obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC   macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC   osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC   arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC   diseases listed in the specification. The probes and antibodies are also
CC   useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC   release disorders, cardiovascular activity disorders, pain perception
CC   disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC   that decrease or increase the expression of galanin receptor (GAL4) can
CC   be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC   nucleic acid is useful for treating the above mentioned disorders by gene
CC   therapy techniques. The present sequence is a novel GPCR polynucleotide
CC   of the invention
XX
SQ   Sequence 1001 BP; 190 A; 315 C; 311 G; 185 T; 0 U; 0 Other;

Query Match             53.8%;  Score 594.4;  DB 6;  Length 1001;
Best Local Similarity   98.7%;  Pred. No. 4.7e-132;
Matches  620;  Conservative  0;  Mismatches  0;  Indels  6;  Gaps  2;

Qy       100 TCTCGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGC 159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 TCTCGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGC 942

Qy       160 CTTACCTGTCAAAAGAGACAGGAGAGATCAACTGCACAGGCATCTCACCTGGTGCACCTGGCT 219
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 CTTACCTGTCAAAAGAGACAGGAGAGATCAACTGCACAGGCATCTCACCTGGTGCACCTGGCT 882

Qy       220 GTGTCGACCTGCTGTTCACCGGTGGCCTTACCGGAAGGTGGTGTGTTATGTGTGGGC 279
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 GTGTCGACCTGCTGTTCACCGGTGGCCTTACCGGAAGGTGGTGTGTTATGTGTGGGC 822

Qy       280 TCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGAC 339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 TCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGAC 762

Qy       340 ACCTACGGGGGGGTCTACTTCATGGCCGTGTGTGAGCGTGGACCATTACCCAGCTGTGGTC 399
             ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 ACCTAC-GGGGGGTCTACTTCATGGCCGTGTGTGAGCGTGGACCATTACCCAGCTGTGGTC 703

Qy       400 TGTGCCCACTGGGGCCCGCCTCCGCACGGCTGGCGCGCCGGCCCAGGCTGGTCTGCGTGGCC 459
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       702 TGTGCCCACTGGGGCCCGCCTCCGCACGGCTGGCGCGCCGGCCCAGGCTGGTCTGCGTGGCC 643

Qy       460 ATCTGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCG 519
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       642 ATCTGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCG 583

Qy       520 CTGGTGGGCAAGCTGCCTGACCTGCCATGAGTACAGCAGCATGAGTGAGTCAGTCCTGGGCTGCCC 579
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       582 CTGGTGGGCAAGCTGCCTGACCTGCCATGAGTACAGCAGCATGAGTGAGTCAGTCCTGGGCTGCCC 523

Qy       580 CTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGATCATCCTGTCC 639
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       522 CTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGATCATCCTGTCC 463

Qy       640 TGCTATATGAAGATCACCTGGAAGCTGTGCAGCACTGGAAGTCTCGGGAGGACACCAGC 699
             ||||||||||||||||||||||||||||||||||||||||||| ||| || | ||||||
Db       462 TGCTATATGAAGATCACCTGGAAGCTGTGCAGCACTGGAAGTCT-GGGAGAACCCAGTGACCAGC 404

Qy       700 AGGAAAGGACCGCCACTGGCGAGGCTGCC 727
             |||||||||| || ||||||||| |||||
Db       403 GGGAAAGGACCACCCGGCGGGGCCAGCC 376
```

SQ Sequence 536 BP; 93 A; 166 C; 182 G; 95 T; 0 U; 0 Other;

Query Match   36.0%;   Score 397.2;   DB 4;   Length 536;
Best Local Similarity   97.5%;   Pred. No. 5.3e-85;
Matches 435;   Conservative   0;   Mismatches   8;   Indels   3;   Gaps   3;

```
Qy  285 CTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGCGGCGTTTGTTGCTCTACACCGACACCTA 344
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8 CTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGCGGCGTTTGTTGCTCTACACCGACACCTA 67

Qy  345 CGGGGGGTCTACCTCATGGCCTCATGTGTGAGCGTGGACCATTACCCAGCTGTCTGTGTGC 404
            ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 C-GGGGGTCTACCTCATGGCCTCATGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGC 126

Qy  405 CCACTGGGGCCCGCGCCTCCGCACGGCTGGCGCGGCCCGCCGCCCGCCAGGCTGGTCTGGCCCATCTG 464
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  127 CCACTGGGGCCCGCGCCTCCGCACGGCTGGCGCGGCCCGCCGCCCGCCAGGCTGGTCTGGCCCATCTG 186

Qy  465 GACCTTGGTCGTCTGCTGCAGACGATGCCCTTGCTTGATGCCATGACCAGCCGCTGGT 524
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  187 GACCTTGGTCGTCTGCTGCAGACGATGCCCTTGCTTGATGCCATGACCAGCCGCTGGT 246

Qy  525 GGGCAAGTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTC-GGGCTGCCCCTCA 583
            |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  247 GGGCAAGTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCCGGGGCTGCCCCTCA 306

Qy  584 TGGTCCTGGTGGCCTTGCCATTGGCTTCTGTGTGGGCCAGTGGGATCATCCTGTCCTGCT 643
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  307 TGGTCCTGGTGGCCTTGCCATTGGCTTCTGTGTGGGCCAGTGGGATCATCCTGTCCTGCT 366

Qy  644 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGCACCCCAGTGACCAGCAGGA 703
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db  367 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT-GGGAGAACCCAGTGACCAGCCAGGGA 425

Qy  704 AAGGACGGCCACTGGCGAGGCTGCCTG 729
            ||||||||||||||||||||||||||
Db  426 AAGGACCACCACCGGCGGGGCAGCCCG 451
```

RESULT 8
ID AAS30791/c
    AAS30791 standard; cDNA; 556 BP.
XX
AC AAS30791;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human cDNA encoding G protein-coupled receptor nGPCR-2343.
XX
KW Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic;
KW cytostatic; cardiant; antidiabetic; anoretic; hypotensive; hypertensive;
KW antiparkinsonian; nootropic; neuroprotective; antidepressant;
KW viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
KW cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
KW obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW atherosclerosis; Parkinson's disease; psychosis; neurological disorder;
KW schizophrenia; migraine; major depression; anxiety; mental disorder;
KW manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.
XX
OS Homo sapiens.
XX
PN WO200166750-A2.
XX
PD 13-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US007322.
XX
PR 08-MAR-2000; 2000US-0187581P.
PR 08-MAR-2000; 2000US-0187582P.
PR 08-MAR-2000; 2000US-0187714P.
PR 08-MAR-2000; 2000US-0187715P.
PR 08-MAR-2000; 2000US-0187825P.
PR 08-MAR-2000; 2000US-0187828P.

PR 08-MAR-2000; 2000US-0187829P.
PR 08-MAR-2000; 2000US-0187830P.
PR 08-MAR-2000; 2000US-0187833P.
PR 08-MAR-2000; 2000US-0187874P.
PR 08-MAR-2000; 2000US-0187928P.
PR 08-MAR-2000; 2000US-0187929P.
PR 08-MAR-2000; 2000US-0187930P.
PR 08-MAR-2000; 2000US-0188049P.
PR 08-MAR-2000; 2000US-0189294P.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS;
XX
DR WPI; 2001-536778/59.
DR P-PSDB; AAU19222.
XX
PT Isolated nucleic acid molecules encoding G protein-coupled receptors
PT termed nGPCR-x, useful in the treatment and diagnosis of viral
PT infections, cancers and mental disorders (e.g. Parkinson's disease and
PT schizophrenia).
XX
PS Claim 4; Page 204-205; 336pp; English.
XX
CC The invention relates to novel isolated nucleic acid molecules encoding G
CC protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,
CC polypeptides, and modulators may be used in the treatment of diseases and
CC conditions such as infections, such as viral infections caused by HIV-1
CC (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and
CC cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC anorexia, hypotension, hypertension, myocardial infarction,
CC atherosclerosis), Parkinson's disease, and psychotic and neurological
CC disorders, including schizophrenia, migraine, major depression, anxiety,
CC mental disorder, manic depression, and dyskinesia, such as Huntington's
CC disease or Tourette's Syndrome and many other diseases and syndromes
CC listed in the specification. nGPCR-x polynucleotides and polypeptides, as
CC well as nGPCR-x modulators, may also be used in diagnostic assays for
CC such diseases or conditions. The present sequence encodes a G protein-
CC coupled receptor of the invention
XX
SQ Sequence 556 BP; 138 A; 139 C; 151 G; 128 T; 0 U; 0 Other;

Query Match   34.3%;   Score 378.4;   DB 4;   Length 556;
Best Local Similarity   98.8%;   Pred. No. 1.7e-80;
Matches 403;   Conservative   0;   Mismatches   1;   Indels   4;   Gaps   2;

```
Qy  697 AGCAGGAAAGGACGACGCCACTGGCGAGGCTGCCTGCTTACGCTGCTGATGCTGCTGATGCTGCTGGCCGTG 756
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  556 AGCAGGAAAGGACGACGCCACTGGCGAGGCTGCCTGCTTACGCTGCTGATGCTGCTGATGCTGCTGGCCGTG 497

Qy  757 GTGGTCGTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGGAGGAGGATGCTC 816
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  496 GTGGTCGTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGGAGGAGGTATGCTC 437

Qy  817 CACCTGCCATCCTGTCGCGAGCGGAGGGCTTCTTACTGTCCTTCAGGCGCCACCGTGGCC 876
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  436 CACCTGCCATCCTGTCGCGAGCGGAGGGCTTCTTACTGTCCTTCAGGCGCCACCGTGGCC 377

Qy  877 CTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTCTTTGCATCCCCATTAC 936
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  376 CTCATGAACATGAACTGTGGCATTACCCC-ATCATTACTTCTTTGCATCCCCATTAC 318

Qy  937 AGGAAATGGCTCCTGGGCATTTTAAAGCTCAAAGGGTCT 996
            ||||||||||||||||||||||||||   |||||||||
Db  317 AGGAAATGGCTCCTGGGCATTTTAAAGCTCAAAGGGTCT---TCCTCCTCCTCCTCCTCC 261

Qy  997 TCCTCCACCCCAGGAAAAGTTCTTCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCT 1056
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 TCCTCCACCCCAGGAAAAGTTCTTCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCT 201

Qy 1057 ATGTTCTTAGCCGGAGCACGTGGTCTAACGTGTCATTTGCTTTATGACT 1104
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  200 ATGTTCTTAGCCGGAGCACGTGGTCTAACGTGTCATTTGCTTTATGACT 153
```

```
CC   nucleic acid is useful for treating the above mentioned disorders by gene
CC   therapy techniques. The present sequence is a novel GPCR polynucleotide
CC   of the invention
XX
SQ   Sequence 330 BP; 46 A; 101 C; 108 G; 75 T; 0 U; 0 Other;

     Query Match     29.7%;  Score 328.4;  DB 6;  Length 330;
     Best Local Similarity   99.7%;  Pred. No. 1.3e-68;
     Matches 329;  Conservative   0;  Mismatches  1;  Indels   0;  Gaps   0;

Qy   349 GGGGTCTACCTCATGGCCTGTGTGGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCAC 408
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGGGTCTACCTCATGGCCTGTGTGGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCAC 60

Qy   409 TGGGGCCCCGGCCTTCCGACAGGCTGGCCGCGCCCAGGCTGGTCGTGGCCATCTGGACC 468
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCGGGCCCCGGCCTTCCGACAGGCTGGCCGCGCCCAGGCTGGTCGTGGCCATCTGGACC 120

Qy   469 TTGGTGCTGCTGCAGACGATGCCCCTTGCTCTTGATGCCCATGACCAGCCGCTGGTGGGC 528
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TTGGTGCTGCTGCAGACGATGCCCCTTGCTCTTGATGCCCATGACCAGCCGCTGGTGGGC 180

Qy   529 AAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCAGTCAGTCCTCGGGGTCGCCCCTCATGGTC 588
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCAGTCAGTCCTCGGGGTCGCCCCTCATGGTC 240

Qy   589 CTGGTGGCCTTTGCCATTGGCTTCGTGTGGGCCCAGTGGGGATCATCCTGCTGCTATATG 648
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTGGTGGCCTTTGCCATTGGCTTCGTGTGGGCCCAGTGGGGATCATCCTGCTGCTATATG 300

Qy   649 AAGATCACCTGGAAGCTGTGCAGCACAGCT 678
         ||||||||||||||||||||||||||||||
Db   301 AAGATCACCTGGAAGCTGGGCAGCACAGCT 330

RESULT 10
AAA30635
ID   AAA30635 standard; cDNA; 1086 BP.
XX
AC   AAA30635;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human G protein-coupled receptor EBI2 cDNA.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening; agonist;
KW   antagonist; ss.
XX
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;  99WO-US023938.
XX
PR   13-OCT-1998;  98US-00170496.
XX
PA   (AREN-) ARENA PHARM. INC.
XX
PI   Behan DP,  Chalmers DT,  Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   P-PSDB; AAY90630.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents.
XX
PS   Example 1; Page 160; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
```

```
RESULT 9
AAS98111
ID   AAS98111 standard; DNA; 330 BP.
XX
AC   AAS98111;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #68.
XX
KW   Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity; diabetes;
KW   hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US015332.
XX
PR   11-MAY-2000; 2000US-0203217P.
PR   18-MAY-2000; 2000US-0205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP,  Miller M,  Burmer G,  Fabre-Suver C,  Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin receptor
PT   polypeptides useful for identifying modulators that are useful for
PT   treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT   stroke.
XX
PS   Claim 2; Page 83; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a nucleic
CC   acid molecule that is at least 80% identical to the G protein-coupled
CC   (GPCR) polynucleotides included in the specification. Also included are
CC   probes based on the GPCR sequences (including antisense probes), a host
CC   cell comprising an expression vector comprising the GPCR sequence,
CC   antibodies raised against the polypeptides, and methods of identifying
CC   modulators of the polypeptides. The polypeptides are useful for
CC   identifying modulator compounds which function as modulators, activators,
CC   repressors, agonists or antagonists of the novel GPCR polypeptides
CC   including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC   described above can be used to detect the presence of the polypeptides
CC   and nucleic acids and are used to diagnose a variety of diseases or
CC   disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC   amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC   carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC   obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC   macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC   osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC   arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC   diseases listed in the specification. The probes and antibodies are also
CC   useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC   release disorders, cardiovascular activity disorders, pain perception
CC   disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC   that decrease or increase the expression of galanin receptor (GAL4) can
CC   be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
```

CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence X-
CC   (AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg or
CC   Ala, and is preferably Lys. When the endogenous residue at this position
CC   is Lys, this residue is replaced by His, Arg or preferably Ala. The 15
CC   amino acid stretch between the substituted amino acid and the Pro may be
CC   endogenous, non-endogenous, or a mixture of endogenous and non-endogenous
CC   residues. The constitutively active GPCRs are useful for identifying
CC   antagonists, agonists and partial agonists for use as pharmaceutical
CC   agents. The mutant proteins are also useful in research settings for
CC   elucidating the roles of the receptors in normal and diseased conditions.
CC   Antagonists for a particular GPCR are useful for treating diseases and
CC   disorders associated with that receptor. Because the novel mutant GPCRs
CC   are constitutively active, they can be used directly for screening of
CC   compounds without the need for endogenous ligands. The present sequence
CC   represents cDNA encoding a human wild-type GPCR used in an
CC   exemplification of the invention. This was cloned and subjected to site-
CC   directed mutagenesis (SDM) to generate DNA encoding the corresponding
CC   mutant of the invention
XX
SQ   Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

Query Match   16.7%; Score 184.8; DB 3; Length 1086;
Best Local Similarity   50.6%;   Pred. No. 4.1e-34;
Matches 475; Conservative   0; Mismatches 457; Indels   6; Gaps   1;

```
Qy  101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
            |||| ||||||||| ||  |||||||||| |  || || ||||||   ||  ||||| | |
Db  104 CTCTGCATTACAGCCTCCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 163

Qy  161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
            | |  ||||||||  |||||| |||||| |  |  |||  |  |  |         |   |
Db  164 TCATTGTTCAAAACAGGAAAAAATCAACTCACCTACCACCCTCTATTCAACAAATTTGGTGA 223

Qy  221 TGTCTGACCTGCTGTTCACCGTGCCTACCGGGAAGGTGGTGTTATGTGCTGGGCT 280
            | |||||  ||  | ||  | || | |  || || | | || |||| ||||||||
Db  224 TTTCTGATATACTTTTTACCACGGCGTTTTGCCTACACGGAATAGCCTACTATGCAATGGGCT 283

Qy  281 CCAGCTGGCCTTTCGGCAAGGGCGGCTCTCGCAGGCGTGACGGCGTTTGTGCTCTCACCGCA 340
            | |||||| | ||| || || | |  |  | ||| | || || |  |||| | |  | |||| |
Db  284 TTGACTGGAATCGGAGAGTGCCTTGTGTGTAGGATAACTGCGCTAGTGTTTTACATCAACA 343

Qy  341 CCTACGGGGGGGGTCTACCTCATGGCCGTGTGTGAGCGTGGACATTACCCAGTGTGTTCT 400
            | ||  | ||| | | || | | |  | | || |||  | || | ||||| ||| |||  |
Db  344 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCCGCTTCATTGCTGTGGTGC 403

Qy  401 GTGCCCACTGGGGCCCGCGCCTCCCGCACGGCTGGCCCAGGCTGGTCTGGCTGGCCA 460
            | ||| ||| ||  |  || ||  || ||  | ||   ||| | | | || |  || |
Db  404 ACCCTCTACGCTACAACAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTG 463

Qy  461 TCTGGACCTTGGTGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC 520
            ||||| | || | | |||| | |||   | || ||| |||| || | | |||||   |  | |
Db  464 TCTGGATTCTCAGTATTGCTCAGAACTCAGAGAAC...TATCAACCCTATGTCAAAGCAGG 523

Qy  521 TGGTGGCAAGCTGGCCTGGCTCATGCCTGGCATGGAGTCAGCCTGCATGGAGTCAGCCTCGGGGCTGCCCC 580
            | | ||| | | | | ||| | ||||| | |||||||| | || | || ||  |    ||| |||| |||||
Db  524 AGGCTGAAAAGGATTACATGCCATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTTCTCCT 583

Qy  581 TCATGGTCCGTGGTGGCCTTTGCCATTGGCTTCTGGGGCCAGTGGGGATCATCCTGTCCT 640
            ||||| || ||| | ||| ||||  || | ||  |||| | || || |  |  | | ||||
Db  584 GGATTCGTGCTTGGGGCATGTTTTCATAGGATATGTGTACTTCCACTTATAAATCATTCTCATCT 643

Qy  641 GCTATTGAGAGATCACCTGGAAGCTGTGCAGCACAGCGTCGGGAGGACACCCAGTGACCAGCA 700
            |||||  | ||||| | | | || ||||||  | |  ||  | ||||||||||| |||| |  |
Db  644 GCTATTCTCAGATCGTCGCAAACTGCCAAACATAACAAACCCCACTCACTGAGA 703

Qy  701 GGAAAGGACGGCCACTGCGGAGGCCGTGCTGGGCCGCTGAGTGGTGGCCGGTGG 760
            | || | ||||||| | |  ||  |  | | ||||| | |||| | | |  |||
Db  704 AATCTGGTGTAAACAAAAGGCTCCAACACAATTATTCTTATTATTGTGTGCGTTGTTC 763

Qy  761 TCTGCTTCAGCCTCCACCACTCAACATCAGCAGTTCATGGCGGAGAGGATGCTCCACC 820
            ||||  || |  ||| || | | | || | ||| | ||| | || || ||| ||| ||
Db  764 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGTTCGTTTCT 823

Qy  821 TGCCATCC------TGTGCCAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
            ||||| ||      |||  || || |   || | | |  | || |  | || ||| |   
Db  824 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTACAGTAT 883

Qy  875 CCCTCATGAACATGAACACTGTGGCATTACCCCAATCATTTACTTCTTTGCATCCACCATT 934
            || |||||||| ||||  | | | ||||| || ||||||| | ||| |||||| | | |||
Db  884 GCCTGATGGAACTTCAATTGCTGCAGTCATGGACCCTTTATCTACTTCTTTGCATGTGAAAGGGT 943

Qy  935 ACAGGAAATGGCTCCTGGCATTTAAAGCTCAAAGGGTCTTCAAAGGGTCTTCCTCCTCCT 994
            | ||| ||| | || ||||||||||||||||||||| |||| ||||| | | || |||| |
Db  944 ATAAGAGAAAGGTTATGAGGATGTCGAAACGGCAAGTGTATGTATGATTTCTAGTGTGCTG 1003

Qy  995 CCTCCCCACCCCAGGAAAAAGCTTCTTCAGAAAACACCA 1032
            || || |||| ||||||||||||||||||||| |||| |
Db 1004 TGAAGTCAGCCCCTGAAGAGAAAATTCACGTGAAATGACA 1041
```

RESULT 11
ACC48996
ID   ACC48996 standard; cDNA; 1086 BP.
XX
AC   ACC48996;
XX
DT   11-AUG-2003 (first entry)
XX
DE   Human G-protein coupled receptor EBI2 coding sequence.
XX
KW   Human; EBI2; G-protein coupled receptor; receptor; neuropeptide;
KW   corticotropin releasing factor; antidepressant; tranquillizer; nootropic;
KW   neuroprotective; cerebroprotective; antiinflammatory; antirheumatic;
KW   antiarthritic; anorectic; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1086
FT   /*tag=   a
FT   /product= "Human EBI2"
XX
PN   EP1286165-A2.
XX
PD   26-FEB-2003.
XX
PF   01-AUG-2002; 2002EP-00255380.
XX
PR   15-AUG-2001; 2001GB-00019920.
XX
PA   (PFIZ) PFIZER LTD.
PA   (PFIZ) PFIZER INC.
XX
PI   Fidock MD, Robas NM;
XX
DR   WPI; 2003-344719/33.
DR   P-PSDB; ABR42051.
XX
PT   Use of urotensin I or sauvagine, or their analogue or mimetic, as ligand
PT   or modulator of EBI2 or for diagnosing or treating depression, anxiety-
PT   related disorders, stress, Alzheimer's disease, stroke, inflammation or
PT   obesity.
XX
PS   Claim 8; Page 22; 29pp; English.
XX
CC   The present sequence is the coding sequence for human G-protein coupled
CC   receptor EBI2. The invention relates to the identification of EBI2 as a
CC   receptor of neuropeptides of the corticotropin releasing factor (CRF)
CC   family. These neuropeptides, e.g. urotensin I or sauvagine (and their
CC   analogues or mimetics), can be used as modulators of EBI2. The invention

```
CC   also provides screening methods to identify agonists and antagonists for
CC   EBI2.Modulators of the receptor may find use in diagnosing or treating
CC   depression, anxiety-related disorders (e.g. panic), stress, Alzheimer's
CC   disease, stroke, inflammatory disorders (e.g. rheumatoid arthritis),
XX   eating disorders (e.g. anorexia) and obesity
SQ   Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

Query Match    16.7%;   Score 184.8; DB 7;  Length 1086;
Best Local Similarity   50.6%;   Pred. No. 4.1e-34;
Matches 475;   Conservative   0;  Mismatches 457;   Indels   6;  Gaps    1;

Qy      101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCCTTTGCC 160
                |||||| || ||| || |||||||||||||||| |||| |||||||||| || || ||
Db      104 CTCTGCATTACAGCCCTGCTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 163

Qy      161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGCCATCCACCCTGGTGCACCTGGCTG 220
Db      164 TCATTGTTCAAAAACAGGAAAAAAATCAACTTCACCCACCCCTCTATTCAACAAATTGGTGA 223

Qy      221 TGTCTGACCTGCTGTTCACCGTGCCCTACCGGGAAGGGTGGTGTGTATGTGCTGGGGCT 280
Db      224 TTTCTGATATACTTTTTACCACGGCCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 283

Qy      281 CCAGCTGCCCTTCGGCAAGGGGCTCTGCAGGCTGACGGCGCCTTTGTGCTCTACACCGACA 340
Db      284 TTGACTGGAGATCGGAAGATGCCTTGTGTAGGATAACTGCCCTAGTGTTTTTACATCAACA 343

Qy      341 CCTACGGGGGGCTACCTCATGGCCTCGTGTGAGCGTGACCATTACCCAGCTGTGGGTCT 400
Db      344 CATATGCAGGTGTGAACTTTATGACCCGCCTGAGTATTGACCGGCTTCATTGCTGTGGTGC 403

Qy      401 GTGCCCACTGGGGCCCGGCCCTCCGCACGGCGTGGCTGGTCGCGCCAGGCCA 460
Db      404 ACCCTCTACGCTACAACAACAGAGATAAAAAGGATTGAACATGGCAAAAGGCGTGTGCATATTTG 463

Qy      461 TCTGGACCTTGGTGCGTGCTGCGAGACGATGCCCTGCTCTTGATGCCCATGACCAAGCCGC 520
Db      464 TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCTCATCAACCTATGTCAAAGCAGG 523

Qy      521 TGGTGGGCAAGCTGGCCTGCATGGAGCTGACAGCAGCAGATCCGAGTCCTCGGGGCTGCCCC 580
Db      524 AGCTGAAAGGATTACATGCATGGAGTATCCATGGAGTGGAAGAAACTAAATCTCTTCCCT 583

Qy      581 TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCT 640
Db      584 GGATTCTGCTTGGGGCATGTTTCATAGGATATGTAGTTCCACTTATAAATCATTCCATCT 643

Qy      641 GCTATATGAAGATCACCTGGAGTCGTGCAGCAGCTCGGGAGGACCCCAGTGACCAGCAGCA 700
Db      644 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 703

Qy      701 GGAAAGGACGCCCACTGGCGAGGCTGCGTCCTACGGTGTCGATGCCGTGGGTG 760
Db      704 AATCTGGTGTAAACAAAAAGGCTCTCAACACCAATTATTCTTCATTATTGTTGTGTTGTTTGGTTC 763

Qy      761 TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGCGGAGAGGGATGCTCCACC 820
Db      764 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 823

Qy      821 TGCCATCC-----TGTTGCCGGAGGGGCTTCTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db      824 CTAATTTCCTGGAATGTAGCCAAAGACATTCCTCCTGCACTTTACAGTAT 883

Qy      875 CCCTCATGAACAATGAACTGTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
Db      884 GCCTGATGAACTTCAATTGCTCGCATGGACCCCTTTTATCTACTTCTTTGCATGTAAAGGGT 943

Qy      935 ACAGGAAATGGCCCTGGGCATTTTAAGCTCAAAGGGTCTTCCCTCCTCCTCCTCCTCCT 994
Db      944 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1003

Qy      995 CCTCCTCCACCCCAGGAGAAAGCTTCTTCAGAAACACCA 1032
```

```
Db         1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041
                 |||| |||| ||| |||| | ||| |||||| ||| ||

RESULT 12
ADC22596:
ID   ADC22596 standard; cDNA; 1086 BP.
XX
AC   ADC22596;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human G protein-coupled receptor cDNA #22.
XX
KW   Human; gene; ss; G protein-coupled receptor; GPCR;
KW   transmembrane-6 region; TM6; intracellular-3 region; IC3.
XX
OS   Homo sapiens.
XX
PN   US6555339-B1.
XX
PD   29-APR-2003.
XX
PF   13-OCT-1998;   98US-00170496.
XX
PR   14-APR-1997;   97US-00839449.
PR   14-APR-1998;   98US-00060188.
PR   26-JUN-1998;   98US-0090783P.
PR   07-AUG-1998;   98US-0095677P.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Liaw CW, Behan DP, Chalmers DT;
XX
DR   WPI; 2003-742861/70.
DR   P-PSDB; ADC22597.
XX
PT   Creating a constitutively active version of an endogenous human G protein
PT   coupled receptor (GPCR) comprises substituting a specific amino acid in
PT   the transmembrane-6 region with a different amino acid, and testing for
PT   constitutive activity.
XX
PS   Example 1; SEQ ID NO 77; 221pp; English.
XX
CC   The invention relates to a method for treating a non-endogenous,
CC   constitutively active version of an endogenous human G protein-coupled
CC   receptor (GPCR) that has a transmembrane-6 (TM6) region and an
CC   intracellular-3 (IC3) region, by substituting a specific amino acid in
CC   the TM6 region with a different amino acid, and testing for constitutive
CC   activity. The method is useful for creating a constitutively active
CC   version of an endogenous human GPCR that comprises a transmembrane 6
CC   region and an intracellular loop 3 region. The altered human GPCR
CC   polypeptides are useful for screening test compounds for identification
CC   of inverse agonists or partial agonists of GPCR polypeptides, which may
CC   have therapeutic uses. The altered GPCRs may also be used in vivo or in
CC   vitro in biological research. A nucleic acid encoding the altered GPCR
CC   may be used to create a transgenic animal expressing the altered GPCR.
CC   The method allows screening for compounds that modulate the activity of a
CC   human G protein-coupled receptor without the need for provision of a
CC   ligand for the receptor. This is particularly useful in allowing
CC   screening of compounds against orphan receptors for which no ligand is
CC   currently known. This sequence represents cDNA encoding a human GPCR
CC   polypeptide of the invention.
XX
SQ   Sequence 1086 BP; 314 A; 242 C; 206 G; 324 T; 0 U; 0 Other;

Query Match    16.7%;   Score 184.8;  DB 9;  Length 1086;
Best Local Similarity   50.6%;   Pred. No. 4.1e-34;
Matches 475;   Conservative   0;  Mismatches 457;   Indels   6;  Gaps    1;

Qy      101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCCTTTGCC 160
                |||||| || ||| || |||||||||||||||| |||| |||||||||| || || ||
Db      104 CTCTGCATTACAGCCCTGCTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 163
```

```
KW   Epstein Barr virus; EBV; induction; detection; diagnosis; lymphocytes;
KW   antigen; growth; differentiation; mediator; infectious mononucleosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            34..1119
FT                  /*tag= a
FT                  /product= "Epstein Barr virus induced polypeptide."
XX
PN   WO9412519-A1.
XX
PD   09-JUN-1994.
XX
PF   08-OCT-1993;  93WO-US009636.
XX
PR   25-NOV-1992;  92US-00980518.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
PI   Birkenbach M,  Kieff E;
XX
DR   WPI; 1994-200183/24.
DR   P-PSDB; AAR54080.
XX
PT   DNA coding for Epstein Barr Virus induced (EBI) polypeptide(s) and
PT   antibodies for EBI1, 2 and 3 - useful for detecting EBV by hybridisation
PT   or by immunoassay.
PS   Claim 2; Page 56-58; 84pp; English.
XX
CC   EBV infected B lymphocytes recapitulate features of antigen stimulation
CC   in enlarging, increasing RNA synthesis, expressing activation antigens
CC   and adhesion molecules, secreting Ig and proliferating. Unlike antigen
CC   stimulated B lymphocytes, EBV infected B lymphocytes continue to
CC   proliferate (in vitro) as immortalised lymphoblastoid cell lines. Because
CC   of the similar effects of EBV and antigen, EBV induced genes are likely
CC   to include mediators of antigen induced B lymphocyte growth or
CC   differentiation. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other;
```

```
Query Match             16.7%;  Score 184.8;  DB 2;  Length 1643;
Best Local Similarity   50.6%;  Pred. No. 4.8e-34;
Matches  475;           Conservative   0;  Mismatches  457;  Indels   6;  Gaps   1;
```

```
Qy      101  CCTCGTTCTACACAGCCTCCTGGTCTTCAGTGCCCTGGAAACATCCTTGCCCTTTGCC 160
                 ||||| |||| ||| ||      ||||  || |        |   |||||| || ||
Db      137  CTCTGCATTACAGCCTCGTCTTCATCATTGGCTGTGGGAAACTTACTTAGCCTTGCGTCG 196

Qy      161  TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATCTCACCTGGTGCACCTGCTG 220
                 ||||| |||| ||| ||   ||||    || |        |   |||||| || ||
Db      197  TCATTGTTCAAAACAGGAAAAAAATCAACTCTTACCGGGAAGGGTGTGTTATGTGCTGGGT 256

Qy      221  TGTCTGACCTGTGTGTCCAGATCTGCCTTCACCGGGAAGGGTGCGTTATGTGCTGGGT 280
                 |||||| |||| |||| |        ||||| |||||||   ||| || | |||||
Db      257  TTTCTGATATACTTTTTACCAGCTTTGCCTACCGTGATGCTACTATGCAATGGGCT 316

Qy      281  CCAGCTGCCTTTCGGCAAGGGGCTCTCGCAGGGGGCTTGTGCTGTACGGCGTTGTTGTGCTGGGT 340
                 |||| ||||    ||||| | ||   ||||    ||||| |         ||||| | |||
Db      317  CCAGCTGCCCTTTCGGCAAGGGGCTCTCGCAGGGGCTGACGCGTTGGCTGGCTAGTGTTTCATCAACA 376

Qy      341  CCTACGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
                 |||||||||   | |||      ||||   |  |||| |  |||||||||||||||||
Db      377  CATATGCAGGTGTGAACTTATGACCCTGAGTATTGACCGCGTCATTGCTTCATTGCTGTGGTGC 436
```

Various additional alignment blocks continue down the page (Qy / Db pairs, positions 401–1041).

```
RESULT 13
AAQ64126
ID   AAQ64126   standard; DNA; 1643 BP.
XX
AC   AAQ64126;
XX
DT   25-MAR-2003  (revised)
DT   03-FEB-1995  (first entry)
XX
DE   Epstein Barr virus induced (EBI-2) gene.
XX
```

PT Assessing or monitoring foetal or placental development - comprises
PT detecting the level or size of Epstein Barr virus induced nucleic acid or
PT protein in maternal serum samples.
XX
PS Example 2; Col 43-46; 45pp; English.
XX
CC The present sequence was used in the development of a novel method for
CC assessing or monitoring foetal or placental development. The method
CC comprises taking a maternal serum sample, and detecting the level or size
CC of Epstein Barr virus (EBV) induced gene or protein 3 (EBI-3) to obtain a
CC result, which can be compared to a control to assess or monitor foetal or
CC placental development
XX
SQ Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other;

Query Match   16.7%;   Score 184.8;   DB 2;   Length 1643;
Best Local Similarity 50.6%;   Pred. No. 4.8e-34;
Matches 475;   Conservative   0;   Mismatches 457;   Indels   6;   Gaps   1;

```
Qy      101 CTCTGTTCTACACAGCCCTCCGTCTTCAGTGCCCTGGGAAACATCCTTGCCCCTTGCC 160
                |||| |||| ||| |||   || |||||||| || || ||| |||  |  || ||||
Db      137 CTCTGCATTACAGCCCTCGTCTTCATCATTGGCGTGCTGGGAAACTTACTAGCCTTGGTCG 196

Qy      161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTCACCTGGCTG 220
               ||||||  | |||| ||||   |  |||||||   |  || |  ||||| |  || ||
Db      197 TCATTGTTCAAAACAGGAAAAAAATCAACTCTACCACCCCTATTCAACAAATTGGTGA 256

Qy      221 TGTCTGACCTGCTGTCACCGGCCTTACCGGAGAGGGTGGTGTTATGTGCTGGGGCT 280
               |||| |||| ||   ||||||||  |||||||||||| |||||||||||   |||||
Db      257 TTTCTGATATACTTTTACCACGGCTTTGCCTACTACTATGCCTACATGATGCAATGGGCT 316

Qy      281 CCAGCTGGCCTTTCGGCAAGGGGCCTCGCAGGTGACGGCGTTTGTGCCTACTACCGACA 340
               |||||| ||| |  |||||  ||||  |||||  | ||| ||| ||||| ||||||||
Db      317 TTGACTCGAGAATCGGAGATCGGCTATCGTGTAGGATAACGTGCCCTAGTGTTTTACATCAACA 376

Qy      341 TTGACTCGAGAATCGGAGATCGGCTATCGTGTAGGATAACGTGCCCTAGTGTTTTACATCAACA 400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      377 CATATGCAGGTGCTCACCTCACGCCTGAGTATTGACCGCTTCATTGCTGGTGTGC 436

Qy      401 GTGCCCACTGGGGCCCGCGCCTCCGCACGGCTGCGCCCAGGCTGGTCTGCTGGCGGCCA 460
               ||| || ||||||||| | | || |  |  |||   |||||  || | |  || || |||
Db      437 ACCCTCACGCTACAACAAGATAAAAAGGATTGAACATGCAAAGGCGTGCATATTTG 496

Qy      461 TCTGGACCTTGGTGCTGCTGGTGGCCCTGTCTCTGATGCCCATGACCAAGCCGCCCGC 520
               || | || ||  |||  | ||   |||||   |   | || |||||||||||||||||
Db      497 TCTGGATTCTAGTAATTGCTCAGACACTCCCATCCTCATCAACCCTATGTCAAAGCAGG 556

Qy      521 TGGTGGGCAAGCTGGCCTGCGTCGATGGAGTACAGCAGCATGGAGCTCAGTCCTCCGGGCTGCCCC 580
               |||| ||||| ||||| | | || |||||||  ||||   ||||| |||  || ||||||| || |
Db      557 AGGCTGAAAGGATTACATGATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 616

Qy      581 TCATGGTCCTGGTGGGCCTTTGCCATTGGCTTCTGTGTGGCCAGTGGGGATCATCCTGTCCT 640
               ||| ||| ||||||  |||  | || ||  ||||  |   ||| ||| || |   |||||||
Db      617 GGATTCTGCTTGGGGCATGTTTCATAGATGAATGTTCATAGAATATGTACTTCCACTTATAAATCAATTCTCATCT 676

Qy      641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGGACACCAGTGACCAGCA 700
               |||| |||||  |||||||||||| ||| |||  |  |||||||||||||  |||||||||
Db      677 GCTATTCCAGATCCGTCTGCGAAACTTCAGAAACTGCCAAACAAAACCCACTCACTGAGA 736

Qy      701 GGAAAAGGACGCCACTGGGCAGGCTGCCCTACGGCTGCTGTGGCCGTGGTGG 760
               || |||  |   |||| |||| |  |  | ||| ||| ||  || |||| ||
Db      737 AATCTGGTGAAACAAAAGACGTCTCAACACAATTATTCTTATTATTGTTGTTCATCT 796

Qy      761 TTCTGCTTCAGCCCCTACCACCTCCAACATCAAGCAGTTCATGGCGAGGGATGCTCCACC 820
               |||| || || || |  |||| || ||| |  ||||||  |||  ||| ||||||  | ||||
Db      797 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 856

Qy      821 TGCCATCC------TGTGCCCAGCCGGAGGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874
               ||| || |      |   || ||  ||| || |   ||    |||  | ||||||||||
Db      857 CTAATTTCCTGAAATGTAGCCAAAGACAATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916

Qy      875 CCCTCATGAACATGGAACTGGTGCATTACCCCAATCATTACTTCTTTGTCATCCACCCATT 934
               ||| ||   ||| |  | ||| |   |  || ||| | || ||||  |  || | ||||
```

Best Local Similarity 50.6%; Pred. No. 4.8e-34;
Matches 475; Conservative 0; Mismatches 457; Indels 6; Gaps 1;

```
Qy    101 CTCTGTTCTACACAGCCCCTCGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
          ||||| ||| || |  || || || ||||||| || || ||||| || || |   |||
Db    137 CTCTGCATTACAGCCTGCTCTTCATCATTGGGCTGCGTGGGAAACTTACTAGCCTTGGTCG 196

Qy    161 TTACCTGTCAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTCGCACCTGGCTG 220
          | || ||| ||| || || ||| || || ||| |   ||| || ||   || ||
Db    197 TCATTGTTCAAAAACAGGAAAAAATCAACTCACTACCACCCTCTATTCAACAAATTTGGTGA 256

Qy    221 TGTCTGACCTGCTGTTCACCGTGCCTACCGGGAAGGGGTGTGTTATGTGCTGGGCT 280
          |     |||| ||       ||   ||  || ||     || ||  || ||||||||
Db    257 TTTCTGATATACTTTTTACCACGGCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 316

Qy    281 CCAGCTGGCCTTTCGGCAAGGGGGCTTCTGCACGGCTGTTGTGCTCTACACCGACA 340
          ||||||||||   |  |||||| | |  || ||| |   ||   ||| || ||   |||
Db    317 TTGACTGAGAATCGGAGAATGCCTTGTGTAGGATAACTGCGCTAGTGTTTTACATCAACA 376

Qy    341 CCTACGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGGTCT 400
          |    |  || | |||   |||   || ||| ||  ||| |   ||| ||  || ||| |
Db    377 CATAATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGCTGTGGTGC 436

Qy    401 GTGCCCACTGGGGCCCGGCCTCCCGCACGGCTGGCCAGGCTGGTCTGGCGTGGCCA 460
          || |  || || || ||   |  ||| || || ||  ||  | || |  || |||
Db    437 ACCCTCTACGCTACAACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG 496

Qy    461 TCTGGACCTTGGTGCTGTCGCTGCAGAGACGATGCCTTGCTTCTTGATGGCCATGACCAGCCGC 520
          ||| |  ||| |||  | ||||   ||    | |  ||| |   ||  || ||| ||| |||
Db    497 TCTGGATTCTAGTATTTGCTCAGAAATGTCTAGAGCTTTGCTCCAACAGCTGGACCAGGTCGG 556

Qy    521 TGGTGGCAAGCTGGCCTGCCATGGAGTACAGCAGGTACAGCATGGAGTCTCGGGCTGCCCC 580
          || ||||  |||| ||  |||  || || ||||| ||   ||| |   ||  ||| |||||
Db    557 AGGCTGAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 616

Qy    581 TCATGGTCTCGTGGCTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCTGTT 640
          |||    | | || || |||||       || || ||| ||   | |  || ||| |||
Db    617 GGATTCTGCTTGGGCCATGTTCATAGGATATGTACTTCCACTTATAATCATTCCTCATCT 676

Qy    641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCAGCTCGGGAGGACCCAGTGACCAGCA 700
          |||  ||| || || |  |||   |||  ||||  ||  |   |  ||||     |  |
Db    677 GCTATTCTCAGATCGTCTGCAAACTCTTCAGAATGTCGCAAACAAAACCCACTCACTGAGA 736

Qy    701 GGAAAGGACGCCACTGGCGAGGCTGCCTGCTTCACGTGCGTGATGCTGGTGGCGTGG 760
          |||   ||| || || | ||  |  || | | |  |  ||   | |  ||  ||| ||
Db    737 AATCTGGTGTAAACAAAAAGCTCTCAACACAATTATTCTTATTATTGTTGTTGTTTGTTC 796

Qy    761 TCTGCTTCAGCCTACCACCTCAACATCAAGCAGTTCATGCGCGAGGATGCTCCACC 820
          ||||| |||| |  |||  | || || || |||| | |   | || |
Db    797 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 856

Qy    821 TGCCATCC------TGTGCCAGCGGAGGGCTTTCTACTGTCCCTTCAGCGCCACCGTGG 874
          |||||||      ||  ||| |  || || |  |  || | || | |||  || |
Db    857 CTAATTTCCTGAAATGTAGCCAAAGACACATTGTTCCAGATTTCTGCACTTTTACAGTAT 916

Qy    875 CCCTCATGAACATGAACTGTGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
          |||||||||| |||| |  ||| ||    | | || || || |||  ||
Db    917 GCCTGATGAACTTCATTGCTCATGGACCCTTTATCTCACTTCTTTGCATGTAAAGGGT 976

Qy    935 ACAGGAAATGGCTCCTGGCATTTAAAGCTTCAAAGGGTCTTCCTCCTCCCTCCTCCCT 994
          || ||||||| || || ||  |||||  || |||| |  || ||  ||  || ||||
Db    977 ATAAGACAAAGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGAGTTCTAGTGCTG 1036

Qy    995 CCTCCTCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
          |  |  |  || | || ||| |  || |||   |||
Db   1037 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1074
```

RESULT 15
ACA56681
ID   ACA56681 standard; cDNA; 1643 BP.
XX
AC   ACA56681;
XX
DT   06-JUN-2003 (first entry)
XX
DE   Human signalling pathway polynucleotide probe SEQ ID NO 1279.
XX
KW   Human; probe; ss; array element; Parkinson's disease;
KW   signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW   immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS   Homo sapiens.
XX
PN   US6500938-B1.
XX
PD   31-DEC-2002.
XX
PF   30-JAN-1998;   98US-00016434.
XX
PR   30-JAN-1998;   98US-00016434.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Au-Young J, Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
XX
PT   Combination of polynucleotide probes, useful as array elements in a
PT   microarray for monitoring the expression of a number of target
PS   polynucleotides.
XX
PS   Claim 1; SEQ ID NO 1279; 65pp; English.
XX
CC   The invention relates to a combination which, comprises a number of
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signalling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html1?DocID=06500938B1
XX
SQ   Sequence 1643 BP; 509 A; 339 C; 285 G; 510 T; 0 U; 0 Other; Length 1643;

Query Match   16.7%; Score 184.8; DB 7; Length 1643;

Search completed: February 21, 2004, 19:27:06
Job time : 494 secs

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      February 21, 2004, 19:12:09 ; Search time 110 Seconds
                                   (without alignments)
                        5569.689 Million cell updates/sec

Title:         US-10-041-615-33
Perfect score: 1104
Sequence:      1 atgataaactaggcccta.........gtgtcattgccttatgact 1104

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued Patents NA:*
            1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES

Result       %   Query
No.   Score Match Length DB ID                     Description
----- ----- ----- ------ -- --                     -----------
   1  184.8  16.7  1086   4 US-09-170-496D-77       Sequence 77, Appl
   2  184.8  16.7  1643   1 US-08-383-750-3         Sequence 3, Appli
   3  184.8  16.7  1643   3 US-08-352-678-3         Sequence 3, Appli
   4  184.8  16.7  1643   4 US-09-536-954-3         Sequence 3, Appli
   5  184.8  16.7  1643   5 US-09-016-434-1279      Sequence 1279, Ap
   6  184.8  16.7  1643   5 PCT-US93-09636-3        Sequence 3, Appl
   7  183.2  16.6  1086   4 US-09-170-496D-205      Sequence 205, App
   8  118.6  10.7  2025   4 US-09-016-434-1482      Sequence 1482, Ap
   9  112   10.1  1020   4 US-09-170-496D-31       Sequence 31, App
  10  112   10.1  1020   4 US-09-170-496D-181      Sequence 181, App
  11  112   10.1  1900   4 US-09-016-434-1484      Sequence 1484, Ap
  12  112   10.1  1901   1 US-08-153-848-43        Sequence 43, Appl
  13  112   10.1  1901   3 US-09-299-843A-43       Sequence 43, Appl
  14  112   10.1  1901   4 US-09-088-337B-43       Sequence 43, Appl
  15  112   10.1  1901   5 PCT-US93-11153-43       Sequence 43, Appl
  16  112   10.1  2453   5 PCT-US95-07180-1        Sequence 1, Appli
  17  106.8   9.7  2218   4 US-09-214-904-3         Sequence 3, Appli
  18  106.8   9.7  2219   4 US-08-432-174A-1        Sequence 1, Appli
  19  106.8   9.7  2272   3 US-08-147-592A-3        Sequence 3, Appli
  20  106.8   9.7  2272   4 US-08-292-594A-3        Sequence 3, Appli
  21  105.4   9.5  1179   4 US-08-148-708-3         Sequence 3, Appli
  22  105.4   9.5  1378   4 US-08-759-848-2         Sequence 2, Appli
  23  105.4   9.5  1378   4 US-08-148-708-6         Sequence 6, Appl
  24  105.4   9.5  1378   4 US-09-016-434-1123      Sequence 1123, Ap
  25  105.4   9.5  1378   1 PCT-US95-09383-2        Sequence 2, Appli
  26  105.2   9.5  1829   2 US-08-411-859-1         Sequence 1, Appli
  27  105.2   9.5  1829   3 US-08-387-707-7         Sequence 7, Appli
```

```
  28         105.2   9.5  1829   4 US-08-405-271A-7        Sequence 7, Appli
  29          99.8   9.0  1098   4 US-09-170-496D-117      Sequence 117, App
  30          99.8   9.0  1098   4 US-09-170-496D-225      Sequence 225, App
  31          99.8   9.0  1174   3 US-09-034-985-1         Sequence 1, Appli
  32          99.8   9.0  1597   2 US-08-724-974A-1        Sequence 1, Appli
  33          99.8   9.0  1697   4 US-09-364-425B-26       Sequence 26, Appl
  34          99.8   9.0  1842   1 US-08-442-134A-1        Sequence 1, Appli
  35          99.8   9.0  1842   1 US-08-444-581B-1        Sequence 1, Appli
  36          99.8   9.0  1842   1 US-08-446-088A-1        Sequence 1, Appli
  37          98.8   8.9   998   4 US-08-432-174A-3        Sequence 9, Appli
  38          98     8.9  1059   4 US-09-576-160B-9        Sequence 9, Appli
  39          97.2   8.8  1773   4 US-09-016-434-1405      Sequence 1405, Ap
  40          95.2   8.6  1062   4 US-09-576-160B-12       Sequence 12, Appl
  41          92.4   8.4  1805   4 US-08-405-271A-18       Sequence 18, Appl
  42          92.4   8.4  1973   4 US-09-016-434-1391      Sequence 1391, Ap
  43          92.4   8.4  1973   4 US-09-023-655-1417      Sequence 1417, Ap
  44          92.4   8.4  3205   4 US-09-976-594-171       Sequence 171, App
  45          91     8.2  1089   4 US-09-170-496D-3        Sequence 3, Appli
```

                              ALIGNMENTS

```
RESULT 1
US-09-170-496D-77
; Sequence 77, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
;  APPLICANT: Behan, Dominic P.
;  APPLICANT: Chalmers, Derek T.
;  APPLICANT: Liaw, Chen H.
;  TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
;  TITLE OF INVENTION: Receptors
;  FILE REFERENCE: AREN-0040
;  CURRENT APPLICATION NUMBER: US/09/170,496D
;  CURRENT FILING DATE: 1998-10-13
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
;  LENGTH: 1086
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-170-496D-77

Query Match             16.7%;  Score 184.8;  DB 4;  Length 1086;
Best Local Similarity   50.6%;  Pred. No. 3.4e-36;
Matches  475;  Conservative   0;  Mismatches  457;  Indels   6;  Gaps     1;

Qy        101 CTCTGTTCTACACAGCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
                   ||||||  ||||| |||| ||||  || |||   |||||  || || || ||||
Db        104 CTCTGCATTACAGCCTCGTCTTCATCATTGGCGTCGTGGAAACTTACTAGCCTTGGTCG 163

Qy        161 TTACCTGTCAAAAGAGGAGGAAGATCAACTGCACAGGCATCTACCTGGTCCACCTGGCTG 220
                    ||||  || ||||||    ||||| |||||| ||||  |  ||
Db        164 TCATTGTCAAAACAGGAAAAAATCAACTCACCACCCCTATTCAACAAATTTGGTGA 223

Qy        221 TGTCTGACCTGCTGTTCACCCTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT 280
                   |||||||| ||| ||  |||  ||||  |||||||||| |||| ||||||
Db        224 TTTCGATATACTTTTTACCACGGCTTTGCCTACAGGCTAGCGAATAGCCTACTATGCAATGGGGCT 283

Qy        281 CCAGCTGGCCTTCGGCAAGGGGCTCTGCAGGCTGCTCGTGTGAGCGTCGTGGACGCGGCGTTTGTGCTCTACACCGACA 340
                   ||||  ||||  |||  |||| ||  |   |||||  | || |||  |||||| |||| ||
Db        284 TTGACTGGAGAATCGGAGAGATCCGTGTGTAGGATAACTGGCTGTGTGTTTTACATCAACA 343

Qy        341 CCTACGGGGGGTCTTACCTCATGGCCGTTATGACCGCTGGACCATTACCCAGCTGTGGTCT 400
                   || || ||| ||| || || |||   ||  |||||  |  || ||   |||| ||
Db        344 CATATGCAGGTGTGAACTTTATGACCTGCCGTATTGACCGGCTTCATTGCTGTGGTGC 403

Qy        401 GTGCCCACTGGGGGCCCCGGCCTCCGCACGGCTGCCGCCCAGGCTGGTCTGCGTGCGGCCA 460
                   |||   |||  ||   ||  ||   |||| ||| ||
Db        404 ACCCTCTACGCTACACAACAAGATAAAAAGGATTGAACATGCAAAAAGGCCGTGTGCATATTTG 463

Qy        461 TCTGGACCTTGGTGCTGCTGCAGACGAGCATGCCCTTGCTCTTGATGCCGTCTTCGATGACCAAGCCGC 520
```

```
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1643 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: both
    TOPOLOGY: linear
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 34..1116
US-08-383-750-3

Query Match     16.7%;   Score 184.8;   DB 1;   Length 1643;
Best Local Similarity  50.6%;           Pred. No. 4e-36;
Matches  475;  Conservative  0;  Mismatches  457;  Indels  6;  Gaps  1;
```

```
Qy  101 CTCTGTTCTCACAGCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
Db  137 CTCTGCATTACAGCCTCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 196
Qy  161 TTACCTGCTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
Db  197 TCATTGTTCAAAACAGGAAAAAAATCAACTCACCACCCCTCATTCAACACAAATTGGTGA 256
Qy  221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGGAAGGGGTGGTGTGTTATGTGCTGGGCT 280
Db  257 TTTCTGATATACTTTTTACCACGGCCTTTGCCTACACGCAATAGCCTACTATGCAATGGGCT 316
Qy  281 CCAGCTGGCCTTCGGCAAGGGGCCTCTGCAAGGGTGCTGCAGGCTCGACGGCGTTGTGCTC 340
Db  317 TTGACTGGAGAATCGGAGACGTCCGTTGTGTAGGAGTAACTGCGCGCAGTGTTTCATCAACA 376
Qy  341 CCTACGGGGGGTCTACCTCATGGCCGTGTTGTGAGCGTGAAGCTGGACCATTACCAGCTGTGGTCT 400
Db  377 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGGCTCATTGCTCATTGCTGTGGTGC 436
Qy  401 GTGCCCACTGGGCCCCGGCCTGGTGGCCAGCTGGTTCGTGGCTGGTGTGCGGCCA 460
Db  437 ACCCTCTACGCTACACAACAGAATAAAAGGATTGAACATGCAAAAGGCGTGTGCATGAAATTG 496
Qy  461 TCTGGACCTTGGTGTGCTGCTGGAGAGATGCGCCTTGCTCTTCATCATCCATCAACGCCGC 520
Db  497 TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG 556
Qy  521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCCTCGGGCTGCCCC 580
Db  557 AGGCTGAAAGGATTACATGCATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 616
Qy  581 TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCGTGTCCT 640
Db  617 GGATTCTGCTTGGGGCATGTTCATAGGATATGTGTACTTCCACTTATAATCATTCTCATCT 676
Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGTCAGCAGCAGCTCGGGAGGACCCAGTGACCAGCA 700
Db  677 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACACCCACTCACTGAGA 736
Qy  701 GGAAAGGACGCCACTGGCCGAGGCTGCCTGCTTACGCTGCTGATGCTGGTGGCCGTGGGTGG 760
Db  737 AATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGTGTGGTTTGTTC 796
Qy  761 TCTGCTTCAGCCCCTACCACCTCCAACATCAAGCAGTTCATGGCGAGGAGGATGCTCCACC 820
Db  797 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 856
Qy  821 TGCCATCC------TGTGCCGAGCGGAGGGGCTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db  857 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTGCACTTTACAGTAT 916
Qy  875 CCCTCATGAACATGAACGTGGCATTACCCCAATCATTACTTCTTCTTGCATCCACCCATT 934
Db  917 GCCTGATGAACTTCAATTGCTCATGCACCCTTTTATCTACTTCTTTGCATGTAAAGGGT 976
Qy  935 ACAGGAAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCCTCCTCCT 994
```

```
Db  464 TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG 523
Qy  521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCCTCGGGCTGCCCC 580
Db  524 AGGCTGAAAGGATTACATGCATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 583
Qy  581 TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCGTGTCCT 640
Db  584 GGATTCTGCTTGGGGCATGTTCATAGGATATGTGTACTTCCACTTATAATCATTCTCATCT 643
Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGTCAGCAGCAGCTCGGGAGGACCCAGTGACCAGCA 700
Db  644 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACACCCACTCACTGAGA 703
Qy  701 GGAAAGGACGCCACTGGCCGAGGCTGCCTGCTTACGCTGCTGATGCTGGTGGCCGTGGGTGG 760
Db  704 AATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGTGTGGTTTGTTC 763
Qy  761 TCTGCTTCAGCCCCTACCACCTCCAACATCAAGCAGTTCATGGCGAGGAGGATGCTCCACC 820
Db  764 TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 823
Qy  821 TGCCATCC------TGTGCCGAGCGGAGGGGCTTCTTACTGTCCCTTCAGGCCACCGTGG 874
Db  824 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTGCACTTTACAGTAT 883
Qy  875 CCCTCATGAACATGAACGTGGCATTACCCCAATCATTACTTCTTCTTGCATCCACCCATT 934
Db  884 GCCTGATGAACTTCAATTGCTCATGCACCCTTTTATCTACTTCTTTGCATGTAAAGGGT 943
Qy  935 ACAGGAAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCCTCCTCCT 994
Db  944 ATAAGAGAAAAGGTTATGAGATGTGAAACGGCAAGTCAGTGTATCGAGTTTCTAGTGCTG 1003
Qy  995 CCTCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCA 1032
Db 1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041
```

RESULT 2

US-08-383-750-3

Db   977 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1036

Qy   995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
               |||||| |||||| ||| ||||
Db   1037 TGAAGTCAGCCCCTGAAGAAATTCACGTGAAATGACA 1074

RESULT 3
US-08-352-678-3
; Sequence 3, Application US/08352678
; Patent No. 6043351
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
;   STREET: 600 Atlantic Avenue
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02210
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/352,678
;   FILING DATE: 30-NOV-1994
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/980,518
;   FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Gates, Edward R.
;   REGISTRATION NUMBER: 31,616
;   REFERENCE/DOCKET NUMBER: B0801/7044
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-720-3500
;   TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1643 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 34..1116
US-08-352-678-3

Query Match   16.7%;   Score 184.8;   DB 3;   Length 1643;
Best Local Similarity   50.6%;   Pred. No. 4e-36;
Matches 475;   Conservative   0;   Mismatches 457;   Indels   6;   Gaps   1;

Qy   101 CTCTGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
               ||||||| ||||| ||||| || ||| |||||||||| ||| || ||| || |
Db   137 CTCTGCATTACAGCCCTCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 196

Qy   161 TTACCTGTCAAAAGAGACAGGAAGATCAACTGCACAGGCATCACCTGGTGCACCTGGCTG 220
               || |||||||||| ||||| | ||||| ||||| |||
Db   197 TCATTGTTCAAAACAGGAAAAAAAATCAACTCACCTCTATTCAACAAATTTGGTGA 256

Qy   221 TGTCTGACCTGCTGTTCACCGTGCCTTACCGTGGGCAGGGTGGTTATGTGCTGGGCT 280
               |||| || |||| |||| || ||| |||| ||| || || || ||
Db   257 TTTCTGATATACTTTTACCACGGCGTTTGCCTACACGGAATAGCCTACTATGGCAATGGGCT 316

Qy   281 CCAGCTGGCCTTTCGGCAAGGGGCGGCTCTGCAGGGTGACGGCGTTTGTGCTCTCACACCGACA 340
               ||||| ||| ||| || ||| || ||| || | || |||| || ||
Db   317 TTGACTGGAGAATCGGAGATGCCTTGTGTGTAGGGATAACTGCGCGTAGTGTTTACATCAACA 376

Db   341 CCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
               ||||| ||| ||| ||| || || |||| | |
Qy   377 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCCTTCATTGCTGTGGTGC 436

Qy   401 GTGCCCACTGGGGCCCGGGCCTTCCGCACACGGCTGCCGCGCCAGCTGGTCTGGCTGGCCA 460
               |||||||||||| | |||||||||| |||||| | || | ||| || ||||
Db   437 ACCCTCACGCTACAACAAGATAAAAGGATTGAACATGCAAAAGGCGTGTGTGCATATTG 496

Qy   461 TCTGGACCTTGGTGCTCGCAGACGATGCCCTTGCTCTTGATGCCCATCAAGCCGC 520
               ||||| ||| |||| |||| ||| ||| ||||||
Db   497 TCTGGATTCTAGTATTTGCTCAGACACTCCCATCAACCCTATGTCAAAGCAGG 556

Qy   521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCAGCATGGAGTCAGTCCTCGGGGCTGCCCC 580
               |||| |||||||| ||||| || |||| ||||| || |||||| || |
Db   557 AGGCTGAAAGGATTACATGCATGGAGTATCAACTTTGAAGAAACTAAAATCTCTTCCCT 616

Qy   581 TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCAGGTCAGTGGGGATCATCCTGTCCT 640
               |||||| ||||||||| |||||| || ||||| || || |||
Db   617 GGATTCTGCTTGGGCATGTTTCATAGAGATATGACTACTTCCACTTATAATCATTCTCATCT 676

Qy   641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGACCAGCA 700
               ||||| |||| ||||||| ||| | | ||| || || ||
Db   677 GCTATTCTCAGATCTGCTGCAAATCTTCAGAACTGCCAAACAAACCCACTCACTGAGA 736

Qy   701 GGAAAGGACGCCACTGGCCAGGGCTGCCTGCTTACGCTGCTGGTGGCCGTGGTGG 760
               || || | || || |||| ||| |||||| ||||||| || || || ||
Db   737 AATCTGGTGTAAACAAAAGGCTCTCAACACAATTATTCTTATTATTGTGTGTTTGTTC 796

Qy   761 TCTGCTTCAGCCCCTACCACCTCAACATCACCAGTTCATCATGGCGAGAGGATGCTCCACC 820
               |||| || |||||||| | | ||| || |||||| ||| ||| | ||| ||
Db   797 TCTGTTTCACCTTACCTACCTACCTTCAATTATTGCAATTATTCAACATATGATTAAGAGAGCTTCGTTTCT 856

Qy   821 TGCCATCC------TGTGCCGAGCGAGGCGTTCTTCACTGTCCCTTCAGGCCACCGTGG 874
               ||| || | || | | | | ||||| | |||
Db   857 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916

Qy   875 CCCTCATGAACAATGAACTGTGGCATTACCCCAATCATTACTTACTTCTTCATCACCACCCATT 934
               |||||||| || ||||| |||| ||| |||| ||| || |||||
Db   917 GCCTGATGAACTTCAATTGCTGCATGGACCCTTTATCTACTTCTTTTGCATGTAAAGGGT 976

Qy   935 ACAGGAAATGCTCCTGGGCCATTTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCT 994
               ||| ||||||| |||| ||||||| ||| ||||
Db   977 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTCAGTGTATCGATTTCTAGTGCTCTG 1036

Qy   995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
               |||||| |||||| ||| ||||
Db   1037 TGAAGTCAGCCCCTGAAGAAATTCACGTGAAATGACA 1074

RESULT 4
US-09-536-954-3
; Sequence 3, Application US/09536954
; Patent No. 6500926
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
;   STREET: 600 Atlantic Avenue
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02210
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/536,954

```
;     FILING DATE:
;     CLASSIFICATION:
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/352,678
;     FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;     NAME: Gates, Edward R.
;     REGISTRATION NUMBER: 31,616
;     REFERENCE/DOCKET NUMBER: B0801/7044
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-720-3500
;     TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 3:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 1643 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
;     FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 34..1116
US-09-536-954-3

Query Match      16.7%;   Score 184.8;  DB 4;  Length 1643;
Best Local Similarity  50.6%;  Pred. No. 4e-36;
Matches  475;  Conservative  0;  Mismatches 457;  Indels  6;  Gaps      1;

Qy    101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    137 CTCTGCATTACAGCCCTCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG 196

Qy    161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    197 TCATTGTTCAAAACAGGAAAAAATCAACTCTACCACCCTCTATTCAACAAATTGGTGA 256

Qy    221 TGTCTGACCTGCTGTTCACCGGGCCTTACCGGGAGAGGGGTGGTGTGTTATGTGCTGGGCT 280
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    257 TTTCTGATATACTTTTTTACCACGGCCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 316

Qy    281 CCAGCTGGCCTTTCGGCAAGGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACA 340
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    317 TTGACTGGAGAATCGGAGATGCCTGTGTGTAGGATAACTGCGCTAGTGTTTACATCAACA 376

Qy    341 CCTACGGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGACCATTACCCAGCTGTGGTCT 400
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    377 CATATGCAGGTGTGAACTTTATGACCGCCTGAGTATTGACCGCTTCATTGCTGGTGGTGC 436

Qy    401 GTGCCCACTGGGGCCCCGGCCCTCCGGCACGGCTGGGCTGGTCTCGCGTGGCGTGGCCA 460
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    437 ACCCTCTACGCGCACAACAAGAGATAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTG 496

Qy    461 TCTGGAACTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC 520
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    497 TCTGGATTCTAGTATTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG 556

Qy    521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTGCCCC 580
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    557 AGGCTGAAAGGATTACATGCCATGCATGGGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 616

Qy    581 TCATGGTCCTGGTGGCCTTTGGCCATTGGCCTTCTGTGGGCCAGTGGGGATCATCCTGTCCT 640
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    617 GGATTCTCGTTGGGGCATGTTTCATAGGATATGTACTTCCACTTATAATCATTCTCATCT 676

Qy    641 GCTATATGAAGATCACCTGGAAGCTGGCAGCAGCAGCTGCGCAGACAGCACCCAGTGACCAGCA 700
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    677 GCTATTCTCAGATCTGCTGCAAACTGTTGCGCAAACTGCCAAACATGCCAAACATCCTTGCCCTTTGCC 160 (?)

Qy    701 GGAAAGGACGCGCCACTGGCGCGAGGCGTGCTGCTTACCGCTGCTGATGCGTGGTCCGTGGTGG 760
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    737 AATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGTTGTGTTGTTC 796

Qy    761 TCTGCTTCAGCCCGTACCACCTCAACATCAAGCAGTTCATGGCGAGGATGCTCCACC 820
```

```
Db    797 TCTGTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 856
              |||| |||||||||||||||||      ||||||
Qy    821 TGTGCCGAGCGGCGGAGGCTTCTTACTGTCCCTTACGTCCCCTTCAGGCCCACCGTGG 874
              |||||||----------||||| ||||||||||
Db    857 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 916

Qy    875 CCCTCATGAACATGAACTGGGCATTACCCCAATCATTCATTTGCATCCAATCATTCTTACTTCTTGC 934
              |||| ||||||||||||||||||| ||||||
Db    917 GCCTGATGAACTTCAATTGCTGTCATGGGTGACCCTTTTATCTACTTCTTTGCATGTAAGGGT 976

Qy    935 ACAGGAAATGGTCCTCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCT 994
              |||||  |||||||||
Db    977 ATAAGAGAAGGTTATGAGGATGCTGAAACGAGCAAGTCAGTGTATCGTATGCGATTTCTAGTGCTG 1036

Qy    995 CCTCCTCCACCCAGGAAAAGCTTCTTCAGAAACACCA 1032
              ||||||| |||||
Db   1037 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1074
```

```
RESULT 5
US-09-016-434-1279
; Sequence 1279, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
;     APPLICANT: Janice Au-Young
;     APPLICANT: Jeffrey J. Seilhamer
;     TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
;     TITLE OF INVENTION: PATHWAY GENE EXPRESSION
;     NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
;     ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;     STREET: 3174 PORTER DRIVE
;     CITY: PALO ALTO
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 94304
; COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/016,434
;     FILING DATE: HEREWITH
;     CLASSIFICATION:
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;     CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;     NAME: Zeller, Karen J.
;     REGISTRATION NUMBER: 37,071
;     REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (650) 855-0555
;     TELEFAX:   (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1279:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 1643 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     IMMEDIATE SOURCE:
;     LIBRARY: GENBANK
;     CLONE: g292056
US-09-016-434-1279

Query Match      16.7%;   Score 184.8;  DB 4;  Length 1643;
Best Local Similarity  50.6%;  Pred. No. 4e-36;
Matches  475;  Conservative  0;  Mismatches 457;  Indels  6;  Gaps      1;

Qy    101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
```

ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRNT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09636
FILING DATE: herewith
CLASSIFICATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1643 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 34..1116
PCT-US93-09636-3

Query Match   16.7%;   Score 184.8;   DB 5;   Length 1643;
Best Local Similarity   50.6%;   Pred. No. 4e-35;   Indels 6;   Gaps 1;
Matches 475;   Conservative   0;   Mismatches 457;

RESULT 6
PCT-US93-09636-3
; Sequence 3, Application PC/TUS9309636
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliot
; TITLE OF INVENTION: Epstein Barr Virus Induced Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

```
Qy  461  TCTGGTGTCTGTGCTGCTGCGACGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC  520
              ||||  ||| |  ||  |  ||  || |  ||||   ||||  |||     ||||||  ||
Db  464  TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG   523

Qy  521  TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGAGTCAGTGTCCTCCGGGCTGCCCC  580
              || | |||| || || ||||| ||||  ||| ||| |  ||  || |||| || || ||
Db  524  AGGCTGAAAAGGATTACAATGCATGGCTGCAGTATCGAAACTTGAAGAAACTAAATCTCTTCCT  583

Qy  581  TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCGTGTGGGCCAGTGGGGGATCATCCTGTCT  640
              ||||| | || || |  ||| |  ||  |  ||| |||  || ||   ||  || ||
Db  584  GGATTCTGCTTGGGGCATGTTTCATAGGATATGTACTTCACTTATAATCATTCTCATCT     643

Qy  641  GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGGACACCCAGTGACCAGCA  700
             ||||  |  ||| | ||||| | || ||| | || |||   ||||| || || ||||  ||
Db  644  GCTATTCTCAGATCGTCTGCGAAACTCCTCAGAATGCCAAACAACAAACCCACTCACTGAGA   703

Qy  701  GGAAAGGACGCCACTGCCGCTGCTTACGCTGCTGATGATGCTGGTGGCCGTGGTGTGGTGG  760
             | |||| | ||  |||  | |  ||   |   ||| ||| |   ||||| |  ||||| |
Db  704  AATCTGGTGTAAACAAAAAGGCTAAAAACAATTCTTATTATTGTTGTGTGTTTGTTC       763

Qy  761  TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTCAGTGCAGAGGGATGCTCCACC     820
              |||| |||| || | |||| ||| ||| | | | ||  || | |||  | ||| |||||
Db  764  TCTGTTTCACACTTACCATGTTGCAATTATTCAACATATGATTATTGAAGAGCTTCGTTTCT  823

Qy  821  TGCCATCC-----TGTGCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGCGCCACCGTGG   874
             ||||| ||        | |  | |  ||| ||  | |||| ||   ||||||||||  ||
Db  824  CTAATTTCCTGGAATGTAGCCAAAGACATTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT  883

Qy  875  CCCTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTACTTCTTTGCATGTAAAGGGT   934
             ||| ||||||| |||| ||| || ||||||||||||||||| ||||||||||||||||| |||
Db  884  GCCTGATGAACTTCAATTGCTGCATGGACCCTTTTATCTACTTCTTTGCATGTAAAGGGT     943

Qy  935  ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCCTCCCTCCTCCT     994
             | || ||| ||| | |   | ||  |||||| || |  || ||  |||  |  | |||
Db  944  ATAAGAGAAAGGTTATGACGATGCTGAAACGGCAAGTCAGTGTATCGCAGTTCTAGTGCTG  1003

Qy  995  CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA  1032
             ||| ||||| || |||| || |  |||||||||||| |
Db 1004  TGAAGTCAGCCCCTGAAGAAATCACGTGAAATGACA    1041
```

```
RESULT 7
US-09-170-496D-205

; Sequence 205, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT:  Behan, Dominic P.
; APPLICANT:  Chalmers, Derek T.
; APPLICANT:  Liaw, Chen W.
; TITLE OF INVENTION:  No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION:  Receptors
; FILE REFERENCE:  AREN-0040
; CURRENT APPLICATION NUMBER:  US/09/170,496D
; CURRENT FILING DATE:  1998-10-13
; NUMBER OF SEQ ID NOS:  294
; SOFTWARE:  PatentIn version 3.1
; SEQ ID NO 205
; LENGTH:  1086
; TYPE:  DNA
; ORGANISM:  Homo sapiens
US-09-170-496D-205

Query Match      16.6%;    Score 183.2;   DB 4;    Length 1086;
Best Local Similarity   50.5%;   Pred. No. 8.4e-36;
Matches  474;  Conservative  0;  Mismatches 458;  Indels  6;  Gaps  1;
```

```
Qy  701  GGAAAGGACGGCCACTGGCGGAGGCGTGCCCTGCTTACGCTGCTGATGCTGCTGGTGTGGTGG  760
             | |||| | || | |   | |  ||   |   ||| ||| |   ||  | ||||| |
Db  737  AATCTGGTGTGTAAACAAAAAGGCTCTCAACACTCAAGCAGTTCATTATTCTTATTATTGTTGTGTTGTTC  796

Qy  761  TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTCAGTGCAGAGGGATGCTCCACC     820
              |||| |||| || | |||| ||| |||| | | | ||  || | |||  | ||| |||||
Db  797  TCTGTTTCACACCATGTTGCAATTATTCAACATATGATTATCTTTGCATGTAAAGGGT     856

Qy  821  TGCCATCC------TGTGCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGCGCCACCATT     874
             ||||| ||        | |  | |  ||| ||  | |||| ||   |||||||||||
Db  857  CTAATTTCCTGGAATGTAGCCAAAGACATTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT   916

Qy  875  CCCTCATGAACATGTGGCATTACCCCAATCATTACTTACTTCTTTGCATGTAAAGGGT   934
             ||| ||||||| | |  | || |||||||||||||||| ||||||||||||||||| ||
Db  917  GCCTGATGAACTTCAATTGCTGCATGGACCCTTTTATCTACTTCTTTGCATGTAAAGGGT    976

Qy  935  ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCCTCCCTCCTCCT     994
             | || ||| ||| | |   | ||  |||||| || |  || ||  |||  |  | |||
Db  977  ATAAGAGAAAGGTTATGACGATGCTGAAACGGCAAGTCAGTGTATCGCAGTTCTAGTGCTG   1036

Qy  995  CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA  1032
             ||| ||||| || |||| || |  |||||||||||| |
Db 1037  TGAAGTCAGCCCCTGAAGAAATCACGTGAAATGACA    1074
```

```
RESULT 8
US-09-016-434-1482

; Sequence 1482, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT:  Janice Au-Young
; APPLICANT:  Jeffrey J. Seilhamer
; TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION:  PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES:  1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
; STREET:  3174 PORTER DRIVE
; CITY:  PALO ALTO
; STATE:  CALIFORNIA
; COUNTRY:  USA
; ZIP:  94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/016,434
; FILING DATE:  HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

```
Qy  101  CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCCTTGCC  160
             ||||| |||| || ||  | ||| | || ||| |  ||| |||||  | | |||  | |
Db  104  CTCTGCATTACAGCCTCGTCTTCATCATTGGCTCTGTGGGAAACTTACTAGCCTTGGTCG  163

Qy  161  TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTGGTGCACCTGGGCTG     220
             | || || ||||  | ||||| | |||| |||  | || ||| | || || |||  |
Db  164  TCATTGTTCAAAAACAGGAAAAATCAACTCAACCCCTATTCAACAAATTGGTGA        223

Qy  221  TGTCTGACCTGCTGCTTCACCGGGAAGGGTGTGTTATGTGCTGTGGGCT            280
             |||| ||  || || || ||| ||| |  | || || | |   ||||||
Db  224  TTTCTGATATACTTTTTACCACGGCTTTGCCTGCAGGAGGGCTTCCTACTATGCAATGGGCT  283

Qy  281  CCAGCTGGCCTTTCGGCAAGGAGGGCTCTGCCAGGCTGACGGCGTTTGTGCTCCTCACACCGACA  340
             ||||| || ||| ||||| | || || | ||   | ||||  ||| |||||||| || ||| |||
Db  284  TTGACTGGAGAATCGGAGATGCCTTGTGTGAGATAACTGCCTAGTGTTTTTACATCAACA     343

Qy  341  CCTACGGGGGGGTCTACCTCATGGCGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT       400
             | || |   | || | |   | | || ||| | || |||| |||||| | ||| | |
Db  344  CATATGCAGGTGAACTTATGACCTGCCTAGTATTGACCGGCTTGCTGTGTGTGGTGC      403

Qy  401  GTGCCCACTGGGCCCGGCGCCTCCGACACGGCCCCAGGCTGGTCTGTCGTGCGGCCA  460
             ||| || |   | ||| || | || | ||  |  || || |||| | | | |||| |
Db  404  ACCCTCTACGCTACAACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTG   463
```

```
;       NAME: Zeller, Karen J.
;       REGISTRATION NUMBER: 37,071
;       REFERENCE/DOCKET NUMBER: PA-0002 US
;       TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (650) 855-0555
;       TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1482:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH: 2025 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       IMMEDIATE SOURCE:
;       LIBRARY: GENBANK
;       CLONE: g984506
; US-09-016-434-1482

Query Match         10.7%;     Score 118.6;  DB 4;   Length 2025;
Best Local Similarity  48.1%;     Pred. No. 8.5e-20;
Matches    399;  Conservative    0;  Mismatches  424;   Indels    6;  Gaps    2;

Qy     95 TGGCGCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCC 154
          || |||| || |||          |||| |||||| | |  |||||||||||||  |||
Db    349 TGCTGCCTGTGTCCTACGGGCGTGGTGTGCGTCTTGGGCTGCTGTGCTCGAACGGCGTGGGCGC 408

Qy    155 TTTGCCTTACCTGTCAAAAGAGCAGGAAGATCAACATCCTTGCACC 214
Db    409 TCTACAATCTTCTTGTGCCGCGTCAAGACCTGGATGCGTCCACGGGAAAGGGTGGTCTGTGTATGTGC 468

Qy    215 TGGCTGCTGTGTCTGACCTGCTGGCCTGGCCTTCACCGTGGCCTCCCTGCCGCTCCCTGCGTC 274
Db    469 TGCTGCTGTGTCTGATGCAGCTGTATGCGCCGCCTCCCTGCGCGTGGTTTTCATTACTACGCCC 528

Qy    275 TGGGCTCCAGCTGGCCTTTCGGCAAGGGCCTCTGCAGGCTGACACGGCGTTTCGTCGCTCACA 334
Db    529 GCGGCGACCACTGGGCCTTCAGCACGGTGCTTCGCTCTCGAAGCTGGTGCGCCTTCCTCTTCTACA 588

Qy    335 CCGACAACTACGGGGGGTCTACTCATGGCGTGTGAGCGTGGACCATTACCCAGCTG 394
Db    589 CCAACCTTTACTGCAGCATCCTCTTCCTCACCGTCCGTGCTCTGGGCG 648

Qy    395 TGGTCTGTGCCCACTGGGCCCTCGGACGGCCTGGCCTCCGCCAGGCTGGTCTGCG 454
Db    649 TCTTACGGACCTCTGCGCTCTGGCGCTCGCCGGCCGGGGGTGGCCG 708

Qy    455 TGGCCATCGGACCTGGACCTTGGTGCTGCTGCGCGAGACGATGCCCTTGCCTCTTGATGCGCCATGACCA 514
Db    709 GGGCCGTGTGGGGGTGTTGGTGCTGGCCTGCCTGCCTGCCTACTTTGTCACCACCA 768

Qy    515 AGCCGCTGGTGGGGCAAGCTGGCCTCATGGAGTACAGCAGCAGAGAGTCAGTCCTCGGGC 574
Db    769 GCGGCGCGGGGGCGCCGCCGCAAGTCGTGCGGCACCATCGC---CGTGGTGCTGCGCTGCGGCT 828

Qy    575 TGCCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCCTTCTGTGGGGCCACTGGGGATCATCC 634
Db    829 TCGTGGCCTACAGCTGCAGTCATGCGGCCTCATGCTCGCGGGTGCCCTTTGCCGTCATCC 888

Qy    635 TGTCCTGCTATATGAAGATCACCTGGAAGTGTCGCAACAGCTCGGGAGGACCCAGTGA 694
Db    889 TTGTCGTTGTTACGTGTCATGGCTCCGGCTAAAGCCAGCCTGACCTTCGGGCG 948

Qy    695 CCAGCAGGAAAGGACGCGCCACTGGCGAGGCTGCCTGCTTACGCTGCTGGTGGCCG 754
Db    949 GCCTCCCTAGGGCCAAGCGCAAGTCGTGCGGCACCATCGC---CGTGGTGCTGCGCTGTCTT 1005

Qy    755 TGGTGTGTCGCTTCAGCCCCTACACCTCTACCCGCCCTACCCGCCCTACCGCGGATGC 814
Db   1006 TCGCCCTCTGCTTCCTGCCATTCGAGTCAGTCATGAGCGAGAGGGGATGC 1065

Qy    815 TCCACCTGCCATCCTGTGCCGAGCCGAGGGCTTTCTTACTGTCCCTTCAGGCCACGTGG 874
Db   1066 TGGACCT---CAGCTGCCACACCTCCAAGCCCATCAACATGATGCGCATCCGG 1122
```

```
Qy    875 CCCTCATGAACATGAACTGGCATTACCCCAATCATTACTTCTTTGC 923
          |||   ||||||||| ||||||| ||| |
Db   1123 CGCTGGCCAGTGCTAACAGTTGCTTGACCCCGTGCTCTCACTTCCTGGC 1171
```

```
Qy
Db

RESULT 9
US-09-170-496D-31
;  Sequence 31, Application US/0917049GD
;  Patent No. 6555339
;  GENERAL INFORMATION:
;  APPLICANT: Behan, Dominic P.
;  APPLICANT: Chalmers, Derek T.
;  APPLICANT: Liaw, Chen W.
;  TITLE OF INVENTION: No. 655339-Endogenous, Constitutively Activated Human G Protein-
;  TITLE OF INVENTION: Receptors
;  FILE REFERENCE: AREN-0040
;  CURRENT APPLICATION NUMBER: US/09/170,496D
;  CURRENT FILING DATE: 1998-10-13
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 31
;       LENGTH: 1020
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;  US-09-170-496D-31

Query Match         10.1%;     Score 112;  DB 4;   Length 1020;
Best Local Similarity  48.9%;     Pred. No. 2.8e-18;
Matches    384;  Conservative    0;  Mismatches  375;   Indels   27;  Gaps    2;

Qy    137 TGGGAAACATCCTTGCCCTTGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAG 196
Db    140 TTGGCAATACCCTGGCTCTGTGGCTTTTCATCCGAGACCACCAAGTCCGGGACCCCGGCCA 199

Qy    197 GCATCTACCTGGTGCACCTGGCTGGTCTGTGTGCTGACCTGCTGTGTCACCGTGGCCTTACCGGGAA 256
Db    200 ACGTGTTCCTGATGCATCTGGCCGGTGCCGACTTGTCGTGCGTGGTCTGCCCACCC 259

Qy    257 GGGTGGTGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGCAAGGGGCTTGCAGGCTGA 316
Db    260 GCCTGGTCTACCACTTCTCTGGGAACCACTGGGCCATTTGGGGAAATCGCATCGCTCA 319

Qy    317 CGGGCGTTTGTCTTCTACACCGACAACTACGGGGGGTCTACCTCATGGCGTGTGAGCG 376
Db    320 CCGGGCTTCCTCTTCTACCTCAACATGTACGGCCAGCATCACTTCCTCACCTGCATCAGCG 379

Qy    377 TGGACCATTACCCAGCTGTGCTGTCTGGCCACTGGGCCCTCGGACGGCTGGCC 436
Db    380 CCGACCGTTTCTCGCCCATTGCACCCGGTCAAGTCCCTGACCTTCTGCTGCTCGAGACGCAGCCTGC 439

Qy    437 GCGCCAGGCTGGTCTGCGTGGCCATCGGACCTTGCTGTGCTGGCCATGGCCCCTGC 496
Db    440 ACGCACACCTGGCCTGCCCTGTCGGCGTGGTGTGGCCTGCAAGCTGGCCATGGCCCCTGC 499

Qy    497 TCTTGATGCCCATGACCAAGCCGCTGGTGGGGCAAGCTGGCCATGGAGTACAGCAGCA 556
Db    500 TGGTGAGCCCACAGACCGTCGAGACCAACCACGGTGGTCTGCGCCTGGTGTACGGGG 559

Qy    557 TGGAGTCAGTCCTCGGGCTGCCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCCTTCTGTG 616
Db    560 AGAAGGC--------CTCCCACCATGCCGCTGTATATGAAGATCACCTGGAAGTGCCTCACCT 607

Qy    617 GGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGGAAGTGTCGCAACAGC 676
Db    608 TCCCGGTTCATCACCACGGTCACCTGTGCTACCTGCTACCTGCTGATCATCCGCGCGGCC 667

Qy    677 CTCGGAGGACCCAGTGACCAGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTTACGGC 736
Db    668 TGCGTGTGAGAGGCCCTCAAGACCAAGGCAGTGCGCATGATCGC--------- 714

Qy    737 TGCTGATGCTGTGGCCGTGGTGGTGGCCTACACCTTCAACATCAAGCAGT 796
Db    715 --ATAGTGCTGGCCATCTTCCTGGTCTTCGTCGTGACCTACCGTCAACCCGTCCCGCT 772
```

Qy　797　TCATGCGAGAGGGATGCTCCACCTGCCATCCTGTCCCAGCGCGGAGGGCTTTCTTACTGT　856

Db　773　ACGTGCTGCACTACCGCAGCCATGGGGCCTCCTGCGCCACCCAGCGCACTCGACCCCTGG　832

Qy　857　CCCTTCAGGCGACCGCACCGTGGCCCTGGCCCTCATGAACATGAACATGAACCCCAATCATTACT　916

Db　833　CAAACCGCATCACCTCCTGCTTCACCAGCCTCAACGGGGCACTCGACCCCATGTATT　892

Qy　917　TCTTTG　922

Db　893　TCTTCG　898

RESULT 10
US-09-170-496D-181
; Sequence 181, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
;　APPLICANT: Behan, Dominic P.
;　APPLICANT: Chalmers, Derek T.
;　APPLICANT: Liaw, Chen W.
;　TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
;　TITLE OF INVENTION: Receptors
;　FILE REFERENCE: AREN-0040
;　CURRENT APPLICATION NUMBER: US/09/170,496D
;　CURRENT FILING DATE: 1998-10-13
;　NUMBER OF SEQ ID NOS: 294
;　SOFTWARE: Patentin version 3.1
; SEQ ID NO 181
;　LENGTH: 1020
;　TYPE: DNA
;　ORGANISM: Homo sapiens
US-09-170-496D-181

Query Match　10.1%;　Score 112;　DB 4;　Length 1020;
Best Local Similarity　48.9%;　Pred. No. 2.8e-18;
Matches　384;　Conservative　0;　Mismatches　375;　Indels　27;　Gaps　2;

Qy　137　TGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGACAG　196

Db　140　TTGGCAATACCCTGGCTCTGTGGCTTTTCATCCGAGACCACCAAGTCCGGGACCCCGGCCA　199

Qy　197　GCATCTACCTCGGTGCACCTGGCTGTGTCTGCAGCTGCTGTTCACCGTGGCCTTACCGGAA　256

Db　200　ACGTGTTCCTGATGCATCCGGCCGGTGGGCCGACTTGTCGTGCGTGGTCCTGCCCACCC　259

Qy　257　GGGTGGTGGTGTTATGTGGGCTCCAGCTGGCCTTCGGCAAGGGGCTCTGCGAGGCTGA　316

Db　260　GCCTGGGCTACCACTTCTCTGGGAACCACCTGGCCATTGGGAAATCGCATGCCGTCTCA　319

Qy　317　CGGCGTTTGTGCTTCACACCGACCTACGGGGGGGTCTACCTCATGGCCTGTGTGGAGCG　376

Db　320　CCGGCTTCCTCTTCTCACCTCAACATGTACGCCAGCGATCTACTTCCTCACCTGCATCAGCG　379

Qy　377　TCGACCATTACCAGCTGTGGTCGTGCCCACTGGGCGCCGCTCCCGCCACGGCCTGGCC　436

Db　380　CCGACCGTTTCTCTGGGCCATTGTGCACCCCGGTCAAGTCCCTCAAGCTCCGCAGCCCCCTCT　439

Qy　437　GGGCCAGGCTGGTCTGCGTGGCCATCTGGACCTTGGTGCTGCTCAGACGATGCCCTTGC　496

Db　440　ACGCACACCTGCCCTGTGCCTTGTGCCTTCCTCTGTGGGTGGTGGTGTGGCCATGGCCCCGCTGC　499

Qy　497　TCTTGATGCCCATGACCAAGCCGCTGGTGGTGGCAAGCTGGCTGCATGGAGTACAGCACGCA　556

Db　500　TGGTGAGCCCACAGACCGTGCAGACCCAACCACACGGTGTTGCGCTGCAGCTGTACCGGGG　559

Qy　557　TGGAGTCAGTCTGCGGCTGCCCTGCCCTCATGGTCCTGCTGCCATTGGCTTCTGTG　616

Db　560　AGAAGGC----------CTCCCACCATGCCCTGTGTCCCTGCCAGTTGGCCTTCACCT　607

Qy　617　GGCCAGTGGGGATCATCCTGCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAG　676

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1901 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 701..1717
US-08-153-848-43

Query Match   10.1%;   Score 112;   DB 1;   Length 1901;
Best Local Similarity   48.9%;   Pred. No. 3.5e-18;
Matches 384;   Conservative   0;   Mismatches 375;   Indels 27;   Gaps 2;

```
Qy    137 TGGGAAACATCCTTGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAG  196
              ||| |||  || |||||  |||||||| || |||   ||||||   |
Db    840 TTGGCAATACCCTGGCTCTGTGGCTTTTCATCCGAGACCACCAAGTCCGGGACCCCCCGCCA  899

Qy    197 GCATCTACCTGGTGCACCTGGCTGTGTCGACCTGCTGTTCACCGTGCCTTACCGGGAA  256
              || ||  ||||| | | |||||  ||||| || |||||||| |||||
Db    900 ACGTGTTCCTGATGCATCTGGCCGTGGCGACTTGTCGTGCGTGTGTCCTGCCCACCC  959

Qy    257 GGGTGGTGGTGTTATGTGCTGGCTTTCGGCAAGGGGCTCTGCAGGCTGA  316
              |||  |||||||||  || |||||||||||  ||   || |
Db    960 GCCTGGTCTACCACTTCTCTGGAACCACTGCCATTTGGGAAATCGCATGCCCGTCTCA  1019

Qy    317 CGGCGTTTGTGCTCTACACCGACACCACCTGGGGGGGTCTACCTCATGGCCTGTGTGAGCG  376
              |||   || |||| |||| ||| ||||| ||| | |||  ||||||||| | | |
Db    1020 CCGGCTTCTTCTTCTACCTCAACATGACGCCAGCATCTACTTCCTCACCTGCATCAGCG  1079

Qy    377 TGGACCATTACCCAGCTGTGTGGTCTGGTGGGCAAGCTGGCCTTCCGACGGCTGGCC  436
              |||||||  ||||||| | ||||||| ||| |||||  ||  | ||| ||||||
Db    1080 CCGACCGTTCCTCGGCCTTCGGGGGCCTTCGGAGCTGGCCTTCCGAGCTCCGAGAGCGATGCCCCTTGC  1139

Qy    437 GCGCCAGGCTGGTCTGCGGTGGCCATCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGC  496
              ||  | ||||| ||||||  | | || ||  |||||| || ||||||| |  ||||||||
Db    1140 AGGCACACCTGGCGTGCCTGTGTGCCTTCCTGTGGCGGTGGCTGTGGCTGTGGCTGCTGGGTGTGGCTGC  1199

Qy    497 TCTTGATGCCCATGACCAAGCGCCTGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCA  556
              |||| ||| | ||| ||  ||||| ||||| || |||||  ||||| |||||||
Db    1200 TGTTGAGCCGCAGAACCGTGCAGACCTGCGCAGACCACCACACGGTGCTGTCTGCTGCGTTTGCCTGCCTCAGTGGTCTTCGGG  1259

Qy    557 TGGAGTCAGTCCTCGGCTGCCTTTGCCATTGGCTTCTGTG  616
              |||||||||||   || |||  | ||||||||||
Db    1260 AGAAGGC----------CTCCCACCATGCCCTGTGTGCCCTGGCCTTTGCCAGTGCCTTTCACCT  1307

Qy    617 GGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGGAAGCTGTGTCAGCACAG  676
              |  | ||| | || |  | | ||  ||||||
Db    1307 TGGGGGGATTGGGGCTTGCTTCCTTGCTGGCCTGTCCTGTGCTGGCTTCACCTTCGCTGCTGCT
```

RESULT 12
US-08-153-848-43
; Sequence 43, Application US/08153848
; Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: 6300 Sears Tower, 233 South Wacker Drive

```
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 701..1717
US-09-299-843A-43

Query Match             10.1%;  Score 112;  DB 3;  Length 1901;
Best Local Similarity   48.9%;  Pred. No. 3.5e-18;
Matches   384;  Conservative   0;  Mismatches 375;  Indels   27;  Gaps    2;

Qy    137 TGGGAAACATCCTTGCCCTTGCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAG 196
          ||| ||   ||  ||| | ||   | || | || || |||| |||  | | ||| |||
Db    840 TTGGCAATACCCTGGCTCTGTGGCTTTTCATCCGAGACCACAGTCCGGGACCCGGCCA 899

Qy    197 GCATCTACCTGTGCACCTGGCTGTGTCACCTGTTCGACCTGGCTTGCCTGGCTTACCTG 256
          |||| | || |||  | ||||| || | |||||| || |||||| | || || | | ||
Db    900 ACGTGTTCCTGATGCATCTGGCCTGGCCGGACTTGTGTGCTGGTCCTGCCCACCCC 959

Qy    257 GGGTGGTGTGTTATGCTGCGGGCTTCCAGCTGGCCTTTCGGCAGGGGCTCTGCAGGGCTGA 316
          |||    || | ||| || ||||| |||||| ||| || || ||| | | ||| | || |
Db    960 GCCTGGTCTACCACTTCTCTGGGAAACCACTGGCCATTTGGGAAATCGCATGCCGTCTCA 1019

Qy    317 CGGGCGTTTGTGCTCTACACCGACACACTACGGGGGGTCTCACCTCATGGCCTGTGTGACCG 376
          ||||| |||| | |  |||  || |||| | || | ||| ||||| ||| ||| |||| ||
Db    1020 CCGGGTTCCTCTTCTACCTCAACATGTACGCCAGCATCTACTTCCTCACCTGCATCAGCG 1079

Qy    377 TGGACCATTACCCAGCTGTGGTCTGTGTGCCCACTGGGGGCCGGCACTGCTGCCC 436
          |||||||| ||    | ||| |   | |||||| || |||| |||| ||| ||||
Db    1080 CCGACCGTTTCTCTGGCCATTGTCACCGGTCAAGTCCCTCCAGCCCCCTCT 1139

Qy    437 GCGCCAAGGCTGCTGGTGCGGTGGCCATCTGGACCTTCGGTGTGCTGCGAGACGATGGCCCTTGC 496
          ||||||   | | || | ||| || ||||| || |  |  ||| || | ||| || || || ||
Db    1140 ACGCACCACCTGGCCTGTGCCTTCCGTGGGTTGGTCGTGTGTGGTGGCCATGGCCCCGCGCTGC 1199

Qy    497 TCTTGATGCCCATGACCAAGCCGCTGGTGTGGGCAAGCTGGCCTGATGAGTACAGCAGCA 556
          |   |  | | |||  |  || | |||| ||| || |||| | ||| |   | | | | |
Db    1200 TGGTGAGCCCACAGACCGTGCAGACCAACCACACGGTGTCTGCGCTGTACCGGG 1259

Qy    557 TGGAGTCAGTCCTCCTCGGGCTGCCCCCTCATGGTCCTTGGCCTTTGCCATTGGCTTGTCTGTG 616
          || |   ||  ||  || |    |||   ||  |  |  || |  |  | || ||  ||| |
Db    1260 AGAAGGC----------CTCCCACCATGCCCTGGTGTCCCTGGTTTCCCGGCAGTGCCTTTCACCT 1307

Qy    617 GGCCAGTGGGGATCATCCTGCCTGCTATATGAAGATCACCTGGAAGTCTGTGCAGCACAG 676
          ||||||||||| |  | |    | || | |||| |||||| |   | || | |||||||
Db    1308 TCCCGTTCATCACCACGGTCACCTGCTACCTGGCGGAAAGGAGCAGTGCGGCGAGGGCC 1367

Qy    677 CTCGGGAGGACCCAGTGACCAGCAGAGGAAAGGACGCCACTGGCCGAGGCGTCCTGCTTACGC 736
          ||||| ||||| |||||| ||     ||| ||| | ||| ||| | || | || | ||||||
Db    1368 TGCGTGTGGAGAAGCGCCCTCAAGACCAAGGCAGTGCCGATGATCGCC---------- 1414

Qy    737 TGTCGATGCTGGTGGCCGTGGTGTCTTCAGCCGTTCTTACTGT 796
          |  | |||| |||| | ||||||  | || || |  |  | | |
Db    1415 --ATAGTGCTGGCCATCTTCCTGTCTTCGTCGTCCTACACCGTCACCGTCCGTCT 1472

Qy    797 TCATGGCGAGAGGGATGCCCACCTGCAGCCGGAGGCGTTTCTTACTGT 856
          ||||  || || | |  ||| ||||| || || || ||| | | ||||
Db    1473 ACGTGCTGCACTACCCCAGCTCACGGGCCTCCTCTGCGCATCCTCGCCCCTGG 1532

Qy    857 CCCTTCAGGCCACCGTGGCCCTCATGAACATGAACTGTGGCATTACCCCAATCATTTACT 916
          || || | ||| | ||| ||   |||| | |||| |  |||| | | || ||| | | ||
Db    1533 CAAACCGCATCACCCTCCTGCCTCACACCGCCTCACGCGGGCACTCACCCCATCATGTATT 1592

Qy    917 TCTTTG 922
          ||||||
Db    1593 TCTTCG 1598
```

RESULT 13
US-09-299-843A-43
;   Sequence 43, Application US/09299843A
;   Patent No. 6107475
;   GENERAL INFORMATION:
;   APPLICANT: Godiska, Ronald
;   APPLICANT: Gray, Patrick W.
;   APPLICANT: Schweikart, Vicki L.
;   TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
;   NUMBER OF SEQUENCES: 66
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
;   ADDRESSEE: Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/299,843A
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 09/088,337
;   FILING DATE: 01-JUN-1998
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/153,848
;   FILING DATE: 17-NOV-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/977,452
;   FILING DATE: 17-NOV-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Jill E. Uhl
;   REGISTRATION NUMBER: 43,213
;   REFERENCE/DOCKET NUMBER: 27866/32059B
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (312) 474-6300
;   TELEFAX: (312) 474-0448
;   TELEX:
;   INFORMATION FOR SEQ ID NO: 43:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1901 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single

RESULT 14
US-09-088-337B-43
;   Sequence 43, Application US/09088337B
;   Patent No. 6348574
;   GENERAL INFORMATION:

```
APPLICANT: Godiska, Ronald
           Gray, Patrick W.
           Schweikart, Vicki L.
TITLE OF INVENTION: No. 6348574el Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
               Borun
    STREET: 6300 Sears Tower, 233 South Wacker Drive
    CITY: Chicago
    STATE: Illinois
    COUNTRY: USA
    ZIP: 60606
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/088,337B
    FILING DATE: 01-Jun-1998
    CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/153,848
    FILING DATE: 17-Nov-1993
    APPLICATION NUMBER: US 07/977,452
    FILING DATE: 17-Nov-1992
ATTORNEY/AGENT INFORMATION:
    NAME: No. 6348574and, Greta E.
    REGISTRATION NUMBER: 35,302
    REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (312) 474-6300
    TELEFAX: (312) 474-0448
    TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 43:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 1901 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
    NAME/KEY: CDS
    LOCATION: 701..1717
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-088-337B-43

Query Match    10.1%;  Score 112; DB 4; Length 1901;
Best Local Similarity 48.9%; Pred. No. 3.5e-18;
Matches 384;  Conservative  0; Mismatches 375; Indels 27; Gaps 2;

Qy   137 TGGGAAACATCCTTGCCCTTGCCTTACCTGTCAAAAGACACGAAGATCAACTGCACAG 196
Db   840 TTGGCAATACCCTGGCTCTGTGGCTTTTCATCCGAGACCACCAAGTCCGGGGACCCCGGCCA 899
Qy   197 GCATCTACCTGGTGCACCTGGCTGTGTCTGACCTGCTGTTCACCGTGCCTTACCGGGAA 256
Db   900 ACGTGTTCCTGATGCATCTGGCCGGCCGACTTGTCCGTGCGTGCTGTCCTGCCACCCC 959
Qy   257 GGGTGGTGTTTATGTGCTGGGCTCCAGCTGCCTTTCGGCAAGGGGCTCTCGAGGCTGA 316
Db   960 GCCTGGTCTACCACTTCTCTGGGAACCACTGGCCATTTGGGAAATCGCATGCCGTCTCA 1019
Qy   317 CGGCCGTTGTGCTCTACACCGACCACCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCG 376
Db  1020 CCGGCTTCCTTCCTTACCTCAACATGTACGCCAGCATCTTCCTCACCTGCATCACGCG 1079
Qy   377 TGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGCGCCCTCGGCGGCCGCACGGCTGCC 436
Db  1080 CCGACCGTTTCCTGGCCATTGTGCACCCGGTCAAGTCCCTCAAGCTCCCGCAGGCCCCTCT 1139
Qy   437 GCGCCAGGCTGGTCTGCGTGGCCATCTGGACCTTGGTCTGCTGCTGCTGGAGACGATGCCCTTGC 496

Db  1140 ACGCACACCTGGCCTGTGCCTTCCTGTCGCTGGGTGGGTGGTGGCGCCATGGGCCCGCGTGC 1199
Qy   497 TCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCCTGCCATGGAGTACAGCAGCA 556
Db  1200 TGGTGAGCCCACAGACCGTGCAGACCAACCACACGGTGTCTGCGTCGCTGCGAGCGTACCGGG 1259
Qy   557 TGGAGTCAGTCCTGGGCTGCCCCTCATGGTCCTGGTGCGTGGCCTTTGCCCATTGGCTTCTGTG 616
Db  1260 AGAGAGGC----------CTCCCACCATGCCCTGGTGTCCCTGGCAGTGAGTCTGCGAGCACAG 1307
Qy   617 GGCCAGTGGGGATCATCCTGCTCCTGCTATATGAAGATCACCTGGAAGCTGTCGAGCACAG 676
Db  1308 TCCCGTTCATCACCACGGTCACCTGCTACCTGCTGATCATCCGAGCAGCCTGGCGAGGGCC 1367
Qy   677 CTCGGGAGGACCCAGTGACCAGCAGGAAAGGACGCCACTGGCCAGGCGTCGCCTGCTTACGC 736
Db  1368 TGCGTGTGGAGAAGCGCCTCAAGACACCAGGCAGTGCCAGTGCCGATGATCGCC--------- 1414
Qy   737 TGCTGATGCTGGTGGCGTGGTGTGTCTGTTCAGCCCCTACCACCTCAACATCAAGCAGT 796
Db  1415 --ATAGTGCTGGCCATCTTCCTGGTCTGCTTCGTGCCTTACCTACCACGTCAACCGCTCCGTCT 1472
Qy   797 TCATGGCGGAGGAGGATGCTCCACCTGCCATCCTGTGCCAGCCGAGGGCTTTCTTACTGT 856
Db  1473 ACGTGTCGCACTACCGCAGCCATGGGCCTTCCTGCCGCCACCAGCCATCCTGGCCCTGG 1532
Qy   857 CCCTTCAGGCCACCTGGCCCTGCTCATGAACATGAACATGTGCAATTACCCCAATCATTACT 916
Db  1533 CAAACCGCATCACCTCCTGCCTCACCAGCCTCAACGGGCCACTCACGGGGCACTCAGCCCCATCATGTATT 1592
Qy   917 TCTTTG 922
Db  1593 TCTTCG 1598

RESULT 15
PCT-US93-11153-43
Sequence 43, Application PC/TUS9311153
GENERAL INFORMATION:
    APPLICANT: Godiska, Ronald
    APPLICANT: Gray, Patrick W.
    APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
    ADDRESSEE: Bicknell
    STREET: 6300 Sears Tower, 233 South Wacker Drive
    CITY: Chicago
    STATE: Illinois
    COUNTRY: USA
    ZIP: 60606
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US93/11153
    FILING DATE:
    CLASSIFICATION:
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/977,452
    FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
    NAME: Noland, Greta E.
    REGISTRATION NUMBER: 35,302
    REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (312) 474-6300
    TELEFAX: (312) 474-0448
    TELEX: 25-3856
```

```
;   INFORMATION FOR SEQ ID NO: 43:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 1901 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 701..1717
PCT-US93-11153-43
```

```
Query Match        10.1%;     Score 112;  DB 5;   Length 1901;
Best Local Similarity  48.9%;  Pred. No. 3.5e-18;
Matches  384;  Conservative  0;  Mismatches  375;  Indels  27;  Gaps  2;
```

```
Qy    137 TGGGAAACATCCTTGCCCTTTGCCTACCTGTCAAAAGAGCAGGAAGATCAACTGACACAG 196
              ||||  ||||| |||  ||  |||| |||||||  |||| |||||  |||| || ||||
Db    840 TTGGCAATACCCTGGTCTGTGGCTTTCATCCGAGACCACACAAGTCCGGGACCCCGGCCA 899

Qy    197 GCATCTACCTGGTGCACCTGGCTGTGTCTGACCTGGCCTGCCTTACCGGGAA 256
              ||  ||||| ||  ||  || |||  |||| | |||| || || ||| |
Db    900 ACGTGTTCCTGATGCATCTGCCCGTGCCGACTTGTCGTGCGTGCCTGCCACCC 959

Qy    257 GGGTGGTGTGTTATGTGCTGGGGTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGA 316
              ||   ||||  || | ||   ||| | |||   ||  |||||||  ||  | |||| |||
Db    960 GCCTGGTCTACCACTTCTCTGGGAACCACTGGCCATTTGGGAAATCGCATGCCGTCTCA 1019

Qy    317 CGGCGTTTGTGCTCTACACCGACACCTACGGCGGGGTCACCTCATGGCCTGTGTGAGCG 376
              ||||||  || |||  ||| || || |  || | |   | | |||||||||| |||
Db   1020 CCGGCTTCCTCTTCTTACCTCAACATGTACGCCAGCATCTACTTCCTCACCTGCATCAGCG 1079

Qy    377 TGGACCATTACCCAGCTGTGGTCGTGCCCACTGGGGCCCGGCCTCCGCACGGCTGGCC 436
              | |||  ||| || ||| | ||   ||||||| || ||  || || ||| || ||
Db   1080 CCGACCGTTTCCTGGCCATTGTGCACCCGGTCAAGTCCCTCAAGTCCGCAGCCCCCTCT 1139

Qy    437 GCGCCAGGCTGGTCTGCGTGGCCATCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGC 496
              || ||  | ||| ||  ||  ||| |  || || ||||| | | || || || ||| ||
Db   1140 ACGCACACCTGCCTGTGGCCTTCCTTCCTGTGGGTGGTCGTGGCCATGGCCCCGCTGC 1199

Qy    497 TCTTGATGCCCATGACCAAGCCGCTGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCA 556
              |||||||  | |    |  |  |  |||||||| || || || | |  ||| |  | |
Db   1200 TGGTGAGCCCACAGACCGTGCAGAGCCAACCAACCACACGGTCCGCGCTGTGTACCGGGG 1259

Qy    557 TGGAGTCAGTGCTCCTCGGGGTCTGCCCCTCATGGTCCTGGTGCCATTGGCTTCTGTG 616
              ||| ||  |||    ||     |||| |||| |||| |  ||||||| |||| ||||
Db   1260 AGAAGGC----------CTCCCACCATGCCCTGGTGTCCTGGCAGTGGCCTTCACCT 1307

Qy    617 GGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGAATCACCTGCAAGCTGCAGCACAG 676
              |||||||  |||| ||||| | |  | ||  | | | |  | || ||| | ||||||
Db   1308 TCCCGTTCATCACCAGGTCACCTGTTACCTGCTGATCATCCGACAGCCTGCGGCAGGGCC 1367

Qy    677 CTCGGGAGGACCCAGTGACCAGGAGCAGGAAAGGACGCCACTGGCGGAGGCTGCCTTACGC 736
              ||||||||| | | ||  | ||   | || | ||| ||  | || ||||| ||| | |||
Db   1368 TGCGTGTGGAGAAGCGCCTCAAGACCAAGGCAGTGCCATGATCGCC---------- 1414

Qy    737 TGCTGATGCTGGTGGGCCGTGGTCTTCTTCAGCCCTACCACCTCAACATCAAGCAGT 796
              ||| |||||| |||  | ||| |  |||  |||| | ||| || || || || |||
Db   1415 --ATAGTGCTGGCCATCTTCCGGTCTCGCTTCGTTGCCCTTACCAGCGTCAACCGCTCCGTCT 1472

Qy    797 TCATGGCGAGAGGGATGCTCCACCTGCCATCCTGTCCGAGCGGAGGGGCTTTCTTACTGT 856
              ||| | | ||||  ||||| |  ||| | || ||| ||  |  | ||| || ||| ||
Db   1473 ACGTGCTGCACTACCCGAGCCAGCCATGGGGCCTCCTGCGCCACCCAGCGCATCCTGCCCTGG 1532

Qy    857 CCCTTCAGGCCACCGTGGCCCTTCATGAACATGAACTGTGGCATTACCCCAATCATTACT 916
              | ||||| |  | ||| || |  ||  |  || | || ||| || || || || ||| ||
Db   1533 CAAACCGCATCACCTCCTGCCTCAACAGCCTCAACGGGCACTGACCCCCATCATGTATT 1592

Qy    917 TCTTTG 922
              ||||| |
Db   1593 TCTTCG 1598
```

```
Search completed: February 21, 2004, 21:28:25
Job time : 115 secs
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 19:18:54 ; Search time 467 Seconds
         (without alignments)
         8277.977 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaactaggccccta.........gtgtcattgctttatgact 1104

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Published Applications_NA:*
    1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
    2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
    3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
    4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
    5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
    6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
    7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
    8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
    9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
   10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
   11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
   12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
   13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
   14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
   15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
   16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
   17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
   18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result NO.	Score	% Query Match	Query Length	DB	Length	ID	Description
1	1104	100.0	1104	15	1104	US-10-041-615-33	Sequence 33, Appl
2	779	70.6	4319	14	1619	US-10-017-161-1619	Sequence 1619, Ap
3	779	70.6	4319	15	1293	US-10-292-798-1293	Sequence 1293, Ap
4	672.8	60.9	1526	9	40	US-09-728-952-40	Sequence 40, Appl
5	397.2	36.0	536	10	35	US-09-791-932-35	Sequence 35, Appl
c 6	378.4	34.3	556	11	64	US-09-805-801-944B-64	Sequence 64, Appl
7	184.8	16.7	1086	14	1	US-10-222-024-1	Sequence 77, Appl
8	184.8	16.7	1086	14	77	US-10-251-385-77	Sequence 77, Appl
9	184.8	16.7	1643	15	1279	US-10-305-720-1279	Sequence 1279, Ap
10	184.8	16.7	2477	14	39	US-10-247-671-39	Sequence 39, Appl
11	183.2	16.6	1086	14	205	US-10-251-385-205	Sequence 205, App
12	183.2	16.6	1638	14	111	US-10-225-567A-111	Sequence 111, App
13	172.2	15.6	1556	15	29	US-10-350-923B-29	Sequence 29, Appl
c 14	148	13.4	594	9	90	US-09-812-102-90	Sequence 90, Appl
c 15	123.4	11.2	548	10	10498	US-09-918-995-10498	Sequence 10498, A
16	119.4	10.8	1020	14		US-10-023-586B-3	Sequence 3, Appli
17	119.4	10.8	1076	16		US-10-333-946-20	Sequence 20, Appl
18	118.6	10.7	2025	14		US-10-225-567A-216	Sequence 216, App
19	118.6	10.7	2025	14		US-10-101-510-722	Sequence 722, App
20	118.6	10.7	2025	15		US-10-305-720-1482	Sequence 1482, Ap
21	117	10.6	2118	14		US-10-101-510-31	Sequence 431, App
22	115	10.4	1017	10		US-09-885-453-3	Sequence 5, Appli
23	115	10.1	1020	14		US-10-079-384-5	Sequence 3, Appli
24	112	10.1	1020	9		US-09-788-133-1	Sequence 1, Appli
25	112	10.1	1020	14		US-10-251-385-31	Sequence 31, Appl
26	112	10.1	1900	14		US-10-251-385-181	Sequence 181, App
27	112	10.1	2070	15		US-10-305-720-1484	Sequence 1484, Ap
28	112	10.1	1082	14		US-10-225-567A-303	Sequence 303, App
29	111.8	10.1	1334	14		US-10-023-586B-1	Sequence 1, Appli
30	111.8	10.1	1334	14		US-10-017-161-707	Sequence 707, App
31	111.8	10.1	3143	15		US-10-292-798-619	Sequence 619, App
32	111.8	10.1	1002	14		US-10-275-910-1	Sequence 1, Appli
33	110.2	10.0	2218	14		US-10-088-726-25	Sequence 25, Appl
34	106.8	9.7	2219	9		US-09-214-904-3	Sequence 3, Appli
35	106.8	9.7	3733	14		US-10-112-599A-1	Sequence 1, Appli
36	106	9.6	3733	14		US-10-005-956-844	Sequence 844, App
37	105.4	9.5	1176	11		US-09-826-509-462	Sequence 462, App
38	105.4	9.5	1378	15		US-10-305-720-1123	Sequence 1123, Ap
39	105.4	9.5	3733	14		US-10-005-956-11	Sequence 11, Appl
40	105.4	9.5	3733	14		US-10-005-956-563	Sequence 563, App
41	105.4	9.5	3733	14		US-10-005-956-565	Sequence 565, App
42	105.4	9.5	3733	14		US-10-005-956-567	Sequence 567, App
43	105.4	9.5	3733	14		US-10-005-956-850	Sequence 850, App
44	105.4	9.5	3733	14		US-10-225-567A-47	Sequence 47, Appl
45	105.2	9.5	1829	9		US-09-823-114-7	Sequence 7, Appli

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-10-041-615-33
; Sequence 33, Application US/10041615
; Publication No. US2004001403BA1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US2004001403BAlel GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 33
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1080)
US-10-041-615-33

Query Match   100.0%; Score 1104; DB 15; Length 1104;
Best Local Similarity 100.0%; Pred. No. 2.6e-310;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGATAAACTAGGCCCTTACTTCACACCACCCACAAAATAAAAACTAAGATTAAAGAT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2
US-10-017-161-1619
; Sequence 1619, Application US/10017161
; Publication No. US2003014366BA1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1619
; LENGTH: 4319
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(4319)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(308)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3128)..(3531)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3660)..(4119)
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(48T
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1619

Query Match   70.6%;   Score 779;   DB 14;   Length 4319;
Best Local Similarity 83.6%;   Pred. No. 1.6e-215;
Matches 1006;   Conservative   0;   Mismatches   0;   Indels 197;   Gaps   3;

```
Qy     97 CGCTCTCGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  156
          ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   2940 CGCTCTCGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 2999

Qy    157 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG  216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3000 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 3059

Qy    217 GCTGTGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGGAAGGGTGGTGTGTTAGTGCTG  276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3060 GCTGTGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGGAAGGGTGGTGTGTTATGTGCTG 3119

Qy    277 GGCTCCAGCTGGCCTTCGGCAAGGGGCTCTGCAGGGCTGACGGCGTTTGTGCTCTCACACC  336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3120 GGCTCCAGCTGGCCTTCGGCAAGGGGCTCTGCAGGGCTGACGGCGTTTGTGCTCTCACACC 3179

Qy    337 GACACCTACGGGGGGGTCACCTCAICGGCCTTCAGCCATTACCCAGCTGTG  396
          ||||||||||         |||||||||||||||||||||||||||||||
Db   3180 GACACCTAC-GGGGGGGTCACCTCAICGGCCTTCAGCCATTACCCAGCTGTG 3238

Qy    397 GTCTGTGCCCACTGGGCGCCGCCTTCGACACGGTCTGGTCTCGCGTG  456
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   3239 GTCTGTGCCCACTGGGCGCCGCCTTCGACACGGTCTGGTCTCGCGTG 3298

Qy    457 GCCATCTGACCTTGGTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG  516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299 GCCATCTGACCTTGGTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG 3358
```

Db   1 ATGATAAAACTAGGCCCTTACTTCACACACCCCACAAAATAAAAACTAAGATTAAAGAT 60

```
Qy     61 TTAAATGTGAAAGCCAAAACTGTGAAAATTTTTTGCGGCTCTCGTTCTACACAGCCCTC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAAATGTGGAAAGCCAAAACTGTGAAAATTTTTTGCGGCTCTCGTTCTACACAGCCCTC 120

Qy    121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGCAGG  180
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTG¹CAAAAGAGCAGG 180

Qy    181 AAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTGTGTCTGACCTGCTGTTCACC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTGTGTCTGACCTGCTGTTCACC 240

Qy    241 GTGGCCTTACCGGGGAAGGGTGGTGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGCAAG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTGGCCTTACCGGGGAAGGGTGGTGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGCAAG 300

Qy    301 GGGCTCTGCAGGCTGACGGCGGGCGTTTGTGCTCTCACACCGACACCTACGGGGGGTCACCTC  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGGCTCTGCAGGCTGACGGCGGGCGTTTGTGCTCTCACACCGACACCTACGGGGGGTCACCTC 360

Qy    361 ATGGGCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGCCCCGCGC  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGGGCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGCCCCGCGC 420

Qy    421 CTCCGCACGGCTGGCCGCGCCGCCGCCAGGCTGGTCTCGCGTGGCCATCTGGACCTTGGTGCTGCTG  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCCGCACGGCTGGCCGCGCCGCCGCCAGGCTGGTCTCGCGTGGCCATCTGGACCTTGGTGCTGCTG 480

Qy    481 CAGAGCATGCCCTTGCTTCGATGCCCATGACCAGCCGCTGGGCAAGCTGGCCTGC  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGAGCATGCCCTTGCTTCGATGCCCATGACCAGCCGCTGGGCAAGCTGGCCTGC 540

Qy    541 ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTCCTCGGGGCTCCCTCATGGTCTCCTGGTGGCCTTT  600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTCCTCGGGGCTCCCTCATGGTCTCCTGGTGGCCTTT 600

Qy    601 GCCATTGGCTTCTGTGTGGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCATTGGCTTCTGTGTGGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG 660

Qy    661 AAGCTGTGCAGCACAGCTCGGGAGGACACCCAGTGACCAGCAGGAAAGACGCCACTGGCGA  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AAGCTGTGCAGCACAGCTCGGGAGGACACCCAGTGACCAGCAGGAAAGACGCCACTGGCGA 720

Qy    721 GGCTGCCTGCTGCTTAGCTGCTGATGCTGTACTGGCCGTGGTCGTGCTTCAGCCCTTACCAC  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GGCTGCCTGCTGCTTAGCTGCTGATGCTGTACTGGCCGTGGTCGTGCTTCAGCCCTTACCAC 780

Qy    781 CTCAACATCAAGCAGTTCATGGCCGAGAGGATGCTCCACCTGCCATCCGTGCGAGCGG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTCAACATCAAGCAGTTCATGGCCGAGAGGATGCTCCACCTGCCATCCGTGCGAGCGG 840

Qy    841 AGGGCTTCTTACTGTCCCTTCAGCCACCGTGCCCTCATGAACATGAACTGTGGCATT  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGGGCTTCTTACTGTCCCTTCAGCCACCGTGCCCTCATGAACATGAACTGTGGCATT 900

Qy    901 ACCCCAATCATTACTTCTTTGCATCACCCATTACAGGAAATGGCTTCTGGGCATTTTA  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACCCCAATCATTACTTCTTTGCATCACCCATTACAGGAAATGGCTTCTGGGCATTTTA 960

Qy    961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCTCCACCCCAGGAGGAAAAGCTTCT  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCTCCACCCCAGGAGGAAAAGCTTCT 1020

Qy   1021 TCAGAAACACCAAGTACACCCAGGCCAGGAGGTTCTATGTTCTTAGGCGAGCACGTGGTC  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TCAGAAACACCAAGTACACCCAGGCCAGGAGGTTCTATGTTCTTAGGCGAGCACGTGGTC 1080

Qy   1081 TAACGTGTCATTTGCTTTATGACT  1104
          ||||||||||||||||||||||||
Db   1081 TAACGTGTCATTTGCTTTATGACT 1104
```

```
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 1293
;   LENGTH: 4319
;   TYPE: DNA
;   ORGANISM: Homo sapiens
; FEATURE:
;   LOCATION: source
; FEATURE:
;   LOCATION: (1)..(4319)
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(308)
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3128)..(3531)
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3660)..(4119)
; FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1)..(48)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1293

Query Match             70.6%;  Score 779;  DB 15;  Length 4319;
Best Local Similarity   83.6%;  Pred. No. 1.6e-215;
Matches 1006;  Conservative    0;  Mismatches    0;  Indels  197;  Gaps    3;
```

```
Qy      517 CGCTCGTGGTGGGCAAGCTGGCCTGCATGAGTACAGCAGCATGGAGTCAGTCCTGCGGGCTG 576
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3359 CGCTGGTGGTGGGCAAGCTGGCCTGCATGAGTACAGCAGCATGGAGTCAGTCCTGCGGGCTG 3418

Qy      577 CCCCTCATGGTCCTGGTGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGATCATCCTG 636
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3419 CCCCTCATGTCCTGGTGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGATCATCCTG 3478

Qy      637 TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT------------ 678
                |||||||||||||||||||||||||||||||||||||
Db     3479 TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACTGGGAGAACCCAGTGACCA 3538

Qy      679 ------------------------------------ 678
Db     3539 GCGGGAAAGGACACCACCCGCGGGGGCAGCCCAGGAGGACCCAGTGACCAACAGAGGAAAGGA 3598

Qy      679 ------------------------------------ 678
Db     3599 CGCCACCGGCGGGGGCAGCCCAGGAGGACCCAGTGACCAGCGGGAAAGGACACCACCCGGCA 3658

Qy      679 ---------------------------------CGG 681
                                                  |||
Db     3659 GGACAGCCCGAGAGACCCAGTGACCAGCGGGAAAGGACGCCACCGGCGGGACAGCCCGG 3718

Qy      682 GAGGACCCAGTGACCAGCAGCAGGAAAGGACCACCAGTGGCGAGGCTGCCTCGCTTACGCTGCTG 741
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3719 GAGGACCCAGTGACCAGCAGCAGGAAAGGACCACCAGTGGCCGAGGCTGCCTCGCTTACGCTGCTG 3778

Qy      742 ATGCTGGTGCCGTGGTGGTCGTTCAGCCCTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG 801
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3779 ATGCTGGTGCCGTGGTGGTCGTTCAGCCCTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG 3838

Qy      802 GCGGAGAGGATGCTTCCACCTGCCATCTCGTGCCGAGCCGAGGGCTTTCTTACTGTCCCTT 861
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3839 GCGGAGAGGGATGCTTCCACCTGCCATCTCGTGCCGAGCCGAGGGCTTTCTTACTGTCCCTT 3898

Qy      862 CAGGGCCACCGTGGCCTCATGAACATGACTGTGCCATTACCCCAATCATTTACTTCTTT 921
                |||||||||||||||||||||||||||||||||  |||||  |||||||||||||
Db     3899 CAGGGCCACCGTGGCCCTCATGAACATGAACTGTGCATTACCCC-ATCATTTACTTCTTT 3957

Qy      922 GCATCCACCATTACAGAGAAATGGCTCCTCGGGCATTTTAAAGTCAAAGGGTCTTCCTCC 981
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3958 GCATCCACCATTACAGAGAAATGGCTCCTCGGGCATTTTAAAGTCAAAGGGTCTTCCTCC 4017

Qy      982 TCCTCCTCCTCCTCCTCCTCCAGGAAAACACCAAGTATCACC 1041
                |||||||||||||||||||||||||||||||||||||||||
Db     4018 TCCTCCTCCTCCTCCTCCTCCAGGAAAACACCAAGTATCACC 4077

Qy     1042 CAGGGCCAGAGGTTCTATGTTCTTAGGCGAGCACGTGTCATTTGCTTTATG 1101
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4078 CAGGGCCAGAGGTTCTATGTTCTTAGGCGAGCACGTGTCATTTGCTTTTATG 4137

Qy     1102 ACT 1104
                |||
Db     4138 ACT 4140
```

```
RESULT 3
US-10-292-798-1293
; Sequence 1293, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT: ASAI, KIYOSHI
;  APPLICANT: AKIYAMA, YUTAKA
;  APPLICANT: ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE: 2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
```

```
Query Match                60.9%;  Score 672.8;  DB 9;  Length 1526;
Best Local Similarity      81.9%;  Pred. No. 7.3e-185;
Matches   901;  Conservative   0;  Mismatches   2;  Indels 197;  Gaps   3;

Qy   98 GCTCTCTGTTCTACACAGCCCTCCTTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 157
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 GCTCTCTGTTCTACACAGCCCTCCTTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 481

Qy  158 GCCTTACCTGTCAAAAGAGCAGAGAGATCAACTGACACAGGCATCTACCTGGTGCACCTGG 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 GCCTTACCTGTCAAAAGAGCAGAGAGATCAACTGACACAGGCATCTACCTGGTGCACCTGG 541

Qy  218 CTGTGTCTGACCTGCTGTTCACGGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGG 277
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  542 CTGTGTCTGACCTGCTGTTCACGGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGG 601

Qy  278 GCTCCAGCTGGCCTTTCGCCAAGGGGCTGCTGAGCGGCGTGACGGGCGTTTGTGCTCTCACCG 337
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  602 GCTCCAGCTGGCCTTTCGCCAAGGGGCTGCTGAGCGGCGTGACGGGCGTTTGTGCTCTCACACCG 661

Qy  338 ACACCTACGGGGGGCTTACCTCATGGCCGTGTGTGAGCGTGGACCATTACCCAGCTGTGG 397
        ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  662 ACACCTAC-GGGGGGCTTACCTCATGGCCGTGTGTGAGCGTGGACCATTACCCAGCTGTGG 720

Qy  398 TCTGTGCCCACTGGGGCCCGGGCCTCCCACGGCCTGGCCGCCCAGCCTGGGTCTGCGTGG 457
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 TCTGTGCCCACTGGGGCCCGGGCCTCCCGCACGGCCTGGCCGCCCAGCCTGGGTCTGCGTGG 780

Qy  458 CCATCTGGACCTTGGTCGTCGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGC 517
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CCATCTGGACCTTGGTCGTCGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGC 840

Qy  518 CGCTGGTGGGCAAGCTGCCTGCCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTGC 577
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 CGCTGGTGGGCAAGCTGCCTGCCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTGC 900

Qy  578 CCCTCATGGTCCTGGTGGGCCTTTGCCATTGGCCATTGGCCTTCTGTGGGCCAGTGGGGATCATCCTGT 637
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 CCCTCATGGTCCTGGTGGGCCTTTGCCATTGGCCATTGGCCTTCTGTGGGCCAGTGGGGATCATCCTGT 960

Qy  638 CCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT------------------------ 678
        ||||||||||||||||||||||||||||||||||||||||
Db  961 CCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCTGGAGAACCCAGTGACCAG 1020

Qy  679 ------------------------------------------------------------ 678
Db 1021 CGGGAAAGGACACCACCCGGCGGGGGAGACCCAGTGACCAGCAGGAGAAGGAC 1080

Qy  679 ------------------------------------------------------------ 678
Db 1081 GCCACCGGCGGGGGCAGCCCCAGGAGGACCCAGGCGGGAAAGGACACCACCAGCAG 1140

Qy  679 -------------------------------------------------------CGGG 682
Db 1141 GACAGCCCGAGAGGACCCAGTGACCAGCAGGAAAGATGCCACCGGCGGGGGACAGCCCGGG 1200

Qy  683 AGGACCCAGTGACCAGCAGGAAAGACGCCACTCGGCGAGGCTGCCTGCTGA 742
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AGGACCCAGTGACCAGCAGGAAAGACGCCACTCGGCGAGGCTGCCTGCTGA 1260

Qy  743 TGCTGGTGGCCGTGGTGTTGCTTCAGCGCTTCAGCCCCTACCACCTTCAACATCAAGCAGTTCATGG 802
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 TGCTGGTGGCCGTGGTGTTGCTTCAGCGCTTCAGCCCCTACCACCTTCAACATCAAGCAGTTCATGG 1320

Qy  803 CGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTCCTTC 862
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 CGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGGAGGGCTTTCTCCTTC 1380

Qy  863 AGGCCACCGTGCCCTCATGAACATGAACTGTGGCATTTTAAAGCTCATTTACTTCTTTG 922
        ||||||||||||||||||||||||||||||||||||
Db 1381 AGGCCACCGTGCCCTCATGAACATGAACTGTGGCATTTTAAAGCTCATTTACTTCTTTG 1439

Qy  923 CATCCACCCATTACAGGAAATGGCTTCCTGGCATTTTAAAGCTCAAAGGGTCTTCCTCCT 982
```

```
Qy  679 ------------------------------------------------------------ 678
Db 3599 CGCCACCGGCGGGGGCAGCCGAGGAGGACCCAGTGACCAGCGGGGAAAGGACACCACCGGCA 3658

Qy  679 ---------------------------------------------------------CGG 681
                                                                  |||
Db 3659 GGACAGCCCGAGAGGACGCCACTCGGCGAGGCTGCCTGCTTACGCTGCTGCGGGGGACAGCCCGG 3718

Qy  682 GAGGACCCAGTGACCAGCAGGACGCCACTCGGCGAGGCTGCCTGCTTACGCTGCTG 741
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3719 GAGGACCCAGTGACCAGCAGGACGCCACTCGGCGAGGCTGCCTGCTTACGCTGCTG 3778

Qy  742 ATGCTGGTGGCCGTGGTGTTCTTGCTTCAGCCCCTTACCACCTCAACATCAAGCAGTTCATG 801
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3779 ATGCTGGTGGCCGTGGTGTTCTTGCTTCAGCCCCTTACCACCTCAACATCAAGCAGTTCATG 3838

Qy  802 GCGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGCGAGGGCTTTCTTTACTTCCCTT 861
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3839 GCGAGAGGGATGCTCCACCTGCCATCCTGTGCCGAGCGCGAGGGCTTTCTTTACTTCCCTT 3898

Qy  862 CAGGCCACCGTGCCCTCATGAACATGAACTGTGGCATTACCCCAATCATTTACTTCTTTT 921
        |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db 3899 CAGGCCACCGTGCCCTCATGAACATGAACTGTGGCATTACCCCA-ATCATTTACTTCTTTT 3957

Qy  922 GCATCCACCATTACAGGAAATGGCTTCCTGGCATTTTAAAGCTCAAAGGGTCTTCCTCC 981
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3958 GCATCCACCATTACAGGAAATGGCTTCCTGGCATTTTAAAGCTCAAAGGGTCTTCCTCC 4017

Qy  982 TCCTCCTCCTCCTCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCAAGTATCACC 1041
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4018 TCCTCCTCCTCCTCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCAAGTATCACC 4077

Qy 1042 CAGGCCAGAGAGTTCTATGTTCTTAGGGGAGCACGTTGGTCTCAACGTGTCATTTGCTTTATG 1101
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 4078 CAGGCCAGAGAGTTCTATGTTCTTAGGGGAGCACGTTGGTCTCAACGTGTCATTTGCTTTATG 4137

Qy 1102 ACT 1104
        |||
Db 4138 ACT 4140
```

RESULT 4
US-09-728-952-40
; Sequence 40, Application US/09728952
; Patent No. US20020111302A1
GENERAL INFORMATION:
; APPLICANT: Tang, Y., Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Dunrui
; APPLICANT: Yamazaki, Vicki
; APPLICANT: Ujwal, Manusha L.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020111302A1e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 799
; CURRENT APPLICATION NUMBER: US/09/728,952
; CURRENT FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 40
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (684)..(1016)
US-09-728-952-40

```
Db    68 C-GGGGGGTCTACCTCATGGCCGCTGTGTGAGCCGGTGGACCATTACCAGCTGTGTCTGTGTGC 126
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   405 CCACTGGGGCCCCGCCTCCGCACGGCTGGCCGGCCGCCAGCCGTCGTCGCGTGGCCATCTG 464
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127 CCACTGGGGCCCCGCCTCCGCACGGCTGGCCGGCCGCCAGCCGTCGTCGCGTGGCCATCTG 186
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   465 GACCTTGGTGCTGCTGCGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGT 524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187 GACCTTGGTGCTGCTGCGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGT 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   525 GGGCAAGTCAGTCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTC-GGGCTGCCCCTCA 583
          |||||||||||  |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   247 GGGCAAGTCAGTCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTGCCCCTCA 306
          |||||||||||  |||||||||||||||||||||||||||||||||||||
Qy   584 TGGTCCTGCGTGGCCTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTCCTGCT 643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 TGGTCCTGCGTGGCCTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCTGCT 366
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   644 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGTCTGGGAGGAGACCCAGTGACCAGCAGGA 703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 ATATGAAGATCACCTGGAAGCTGTGCAGCACAGTCTGGGAGGAGACCCAGTGACCAGCGGGA 425
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   704 AAGGACGCCACTGGCGCAGAGCCTGCCTG 729
          ||||||||||||||||||||||||||
Db   426 AAGGACACCACCGGCGCGGCGGCAGCCCG 451
```

```
Query Match             34.3%;  Score 378.4;  DB 11;  Length 556;
Best Local Similarity   98.8%;  Pred. No. 1.8e-99;
Matches   403; Conservative   0; Mismatches   1; Indels  4;  Gaps  2;

Qy   697 AGCAGGAAAGGACGCCACTGGCGAGGCTGCCTACCGTGCTGATGCTGCTGGCCGTG 756
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1440 CATCCACCCATTACAGGAGAAATGGCTCCTGGGCATTTTAAAGCTCAAAGGGTCTTCCTCCT 1499
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   983 CCTCCTCCTCCTCCCTCCTCC 1002
         ||||||||||||||||| ||
Db  1500 CCTCCTCCTCCTCCTCCACC 1519
```

```
Query Match             36.0%;  Score 397.2;  DB 10;  Length 536;
Best Local Similarity   97.5%;  Pred. No. 6e-105;
Matches   435; Conservative   0; Mismatches   8; Indels  3;  Gaps  3;

Qy   285 CTGGCCTTCGGCAAGGGGCTCTGCAGGCTGACGGGCGCTTGTGCTCTTACACCGACACCTA 344
         || |||||| ||||||||||  ||||||||||||||||||||| |||||||||||||||||
Db     8 CTGGCCCTTCGGCAAGGGGCTCTGCAGGCTGACGGGCGCTGAGCGTTTGTGCTCTTACACCGACACCTA 67

Qy   345 CGGGGGGGTCTACCTCATGGCCGCTGTGTGAGCGTGATGCTGCTGGCCGTG 404
         |  ||||||||||||||||||||||||||||||
```

Db   556 AGCAGGAAAGGACGCCACTGGCGAGGCTGCCTGCTTACGGCTGCGCTGATGCTGGTGGCCGTG 497
         ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Qy   757 GTGGTTGCTTGCTTCAGCCCTACCACCTCAACATCAAGCAGTTCATGGCGGAGGAGGATGCTC 816

Qy   496 GTGGGTCGTTCAGCCCTACCACCTCAACATCAACATCAAGCAGTTCATGGCCGAGAGTATGCTC 437
         |||||||||||||||||||||||||||||||||||||||||||
Db   817 CACCTGCCATCCTGTGCCGAGCGGAGGGCTTCTTACTGTCCCTTCAGGCCACCGTGGCC 876

Db   436 CACCTGCCATCCTGTGCCGAGCGGAGGGCTTCTTACTGTCCCCTTCAGGCCACCGTGGCC 377
         |||||||||||||||||||||||||||||||||||
Qy   877 CTCATGAACATGAACTGTGGCATTACCCCCAATCATTACTTCTTTGCATCCACCCATTAC 936

Qy   376 CTCATGAACATGAACTGTGGCATTACCCC-ATCATTACTTCTTTGCATCCACCCATTAC 318
         |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   937 AGGAAATGGCTCCTGGCATTTAAAGCTCAAAGGGTCTTCCTCTCCTCCTCCTCCCC 996

Db   317 AGGAAATGGCTCCTGGCATTTAAAGCTCAAAGGGTCT---TCCTCCTCCTCCTCCC 261
         ||||||||||||||||||||||||||||||||||| |||||||||||||||
Qy   997 TCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCCAAGTATCACCCAGGCCAGAGGTTCT 1056

Qy   260 TCCTCCACCCCAGGAAAAAGCTTCTTCAGAAACACCCAAGTATCACCCAGGCCAGAGGTTCT 201
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1057 ATGTTCTTAGGGGAGCACGTGGTCTAAGGTGTCATTTGCTTTATGACT 1104

Db   200 ATGTTCTTAGGGGAGCACGTGGTCTAAGGTGTCATTTGCTTTATGACT 153

RESULT 7
US-10-222-024-1
; Sequence 1, Application US/10222024
; Publication No. US20030104487A1
; GENERAL INFORMATION:
;   APPLICANT: Pfizer Ltd. (GB)
;   APPLICANT: Pfizer Inc. (EP except GB, US, JP)
;   TITLE OF INVENTION: Neuropeptide receptor and uses thereof
;   FILE REFERENCE: PCS22032
;   CURRENT APPLICATION NUMBER: US/10/222,024
;   CURRENT FILING DATE: 2002-08-15
;   PRIOR APPLICATION NUMBER: GB 0119920.7
;   PRIOR FILING DATE: 2001-08-15
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-222-024-1

Query Match   16.7%; Score 184.8; DB 14; Length 1086;
Best Local Similarity   50.6%; Pred. No. 4.8e-43;
Matches   475; Conservative   0; Mismatches 457; Indels   6; Gaps   1;

Qy   101 CTCTGTTCTACACAGCCCTCCTGGTTCTTCAGTGCCCTGGAAACATCCTTGCCCTTTGCC 160
         |||||||||||||||| |||||||||| || |||| || || || ||||| |||| |||
Db   104 CTCTGCATTACAGCCTGCTTTGGGCTTCGTCGTTCATCATTGGGCTCATTGGCTTGGTCG 163

Qy   161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
         || |||| ||| ||||| || ||||| || |||| ||||| ||||| || ||
Db   164 TCATTGTTCAAAACAGGAAAACAGGAAAAAAATCAACTCTACCACCCCTCTATTCAACAAAATTTGGTGA 223

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT 280
         | ||| || |||| || ||| ||| || |||||| ||| | | | ||| |||
Db   224 TTTCTGATATATCTTTTTACCACGGCGCTTTTTACCACGCTTTGCCTAGCCTACTATGCCGAATAGGCTGCAATGGGCT 283

Qy   281 CCAGCTGGCCTTCGCGCAAGGGGGCTCTGCAGCGGCGGCTGACGCCTGTTGTGCCTCTACACCGACA 340
         |||| ||| | || ||| || | || | || | || | | || || | |||| ||
Db   284 TTGACTGAGGATCGGAGAATCGGAGAGATCGGAGGATGTGTGTGTAGGGATAACTGCGCTTCATTGCGCGGTGCTAGTGTTTTACATCAACA 343

Qy   341 CCTACGGGGGGTCTACCTCATGCCATCATGTGAGCCGTGGACCATTACCCAGTCTGTGGTCT 400
         ||| || | || || || ||||| | ||| |||||| |||| || || |||
Db   344 CATATGCAGGTGTGAACTTTATGACCTTGCCTGAGTATTGACCCGCTTCATTGCTTGTGGTGC 403

Qy   401 GTGCCCACTGGCGCCCGGCCCCGGCCCCGGCCGGCTCCGCACGCGCGCCGCCGCAGGCTGCTGGTCGCGCTGGCCA 460
         || |||| ||| | | || || |||| | || | || | |||| | |||||||| || || |||| ||
Db   101 CTCTGTTCTACACAGCCCTCCTGGTTCTTCAGTGCCCTGGAAACATCCTTGCCCTTTGCC 160

Db   404 ACCCTCTAGGCTACACACAGATAAAAGGATTGAACATGCAAAAAGGCGGTGCAATATTTG 463
         | ||| || ||| || || || ||||| |||| ||| ||| || |||||||||
Qy   461 TCTGGACCTTGGTGCTGCTGCCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC 520

Qy   464 TCTGGATTCCAGTATTTGCTCAGACACTCCCATCATCCTCATCAACCTCATATGTCAAAGCAGG 523
         ||||| | |||| | || || ||| ||||| ||| || | | | | || || |||
Db   521 TGGTGGCAAGCTGCCTGGCTGCATGAGTACAGCAGCATGGAGTCAGTCCTCCGGGCTGCCCC 580

Db   524 AGGCTGAAAGGATTACATGCATGGAGTATCCCAAACTTTGAAGAAACTAAATCTTTCCCT 583
         ||||||| | || ||||| || ||||| | | | || || ||| ||||| |||
Qy   581 TCATGTCCTCGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCT 640

Qy   584 GGATTCTGCTTGGGCATGTTCATAGGATATGTACTTCCACTTATAATCATTCTCATCT 643
         ||||||||||| || ||||| || ||| || | || || ||| | |||| | ||
Db   641 GCTATATGAAGATCACCTGGAAGTGTGCAGCACAGCTTCGGGAGGACACCCAGTGACCAGCA 700

Db   644 GCTATTCTCGACTCGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 703
         |||||||| | |||||| || ||||||| | ||| || ||||||||| | |||| |
Qy   701 GGAAAGGACGGCACTGGCGAGGCTGCCTGCTTACGGCTGCTGTGATGGTCGGTGGTGG 760

Qy   704 AATCCTGGTGTAAACAAAAAGGCTCTCAACACCTCAACATCAAGCAGTTCATGGCGAGAGGATGCTCACCACC 763
         ||| |||| ||| | |||||| || | || ||||| ||| || ||||| | || || || || | |
Db   761 TCTGCTTCAGCCCCTACCACCTCCAACATCAAGCAGTTCATGGCGAGAGGGATGCTCCACC 820

Db   764 TCTGTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAGACTTCGTTTCT 823
         |||||| || || || ||| | || || ||||| || | ||| | || | |
Qy   821 TGCCATCC------TGTGCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGG 874

Qy   824 CTAATTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 883
         ||||||||||| |||| |||| |||| | ||| || || |||||||| || ||| |
Db   875 CCCTCATGAACATGAACTGTGGCATTACCCCCAATCATTACTTCTTTGCATCCACCCATT 934

Db   884 GCCTGATGAACTTCAATTGCTGCATGGACCACCCTTTATCATTCTTTGCATGTAAAGGGT 943
         || || | || || ||| || |||| || | || | ||||| |||||| |||||| |
Qy   935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCT 994

Qy   944 ATAAGAGAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGTG 1003
         || ||||| | || || ||||| |||| || ||||| | ||||| ||||| |
Db   995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032

Db 1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041

RESULT 8
US-10-251-385-77
; Sequence 77, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US2003010529231-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/10/251,385
;   CURRENT FILING DATE: 2002-09-20
;   PRIOR APPLICATION NUMBER: US/09/170,496
;   PRIOR FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: Patentin version 3.1
; SEQ ID NO 77
;   LENGTH: 1086
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-251-385-77

Query Match   16.7%; Score 184.8; DB 14; Length 1086;
Best Local Similarity   50.6%; Pred. No. 4.8e-43;
Matches   475; Conservative   0; Mismatches 457; Indels   6; Gaps   1;

Qy   101 CTCTGTTCTACACAGCCCTCCTGGTTCTTCAGTGCCCTGGAAACATCCTTGCCCTTTGCC 160

```
;   CURRENT FILING DATE: 2002-11-26
;   PRIOR APPLICATION NUMBER: 09/016,434
;   PRIOR FILING DATE: 1998-01-30
;   NUMBER OF SEQ ID NOS: 1490
;   SOFTWARE: PERL Program
;   SEQ ID NO 1279
;   LENGTH: 1643
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   OTHER INFORMATION: GenBank ID No. US20040010136A1 g292056
US-10-305-720-1279

Query Match             16.7%;   Score 184.8;  DB 15;  Length 1643;
Best Local Similarity   50.6%;
Matches  475;   Conservative    0;   Mismatches  457;   Indels    6;   Gaps    1;
                Pred. No. 5.8e-43;
```

```
Qy    101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
              |||  |||| |||||| ||| |||||||||| ||   ||||||  |||||| |
Db    137 CTCTGCATTACAGCCCTCGTCTTCATCATTGGGCTCGTCGTGGGAAACTTACTAGCCTTGGTCG 196

Qy    161 TTACCTGTCAAAAGAGCAGGAAGATCAACTCACCTGGTCGACCTGCACCTGGCTG 220
              || |||||||   ||  ||| ||||| |  |||| |||    || |
Db    197 TCATTGTTCAAAACAGGAAAAAAATCAACTCTACCACCCTCTATTCAACAAATTTGGTGA 256

Qy    221 TGTCTGACCTGCTGTTCACCGGTGCCTTACCGGGAAGGGTGTGTGTTATGTGCTGGGCT 280
              |||||||  |  ||| |  ||||   |||||| |   || | ||
Db    257 TTTCTGATATACTTTTTACCACGGCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 316

Qy    281 CCAGCTGGCCTTTCGGCAAGGAGCTCTGCAGGGTGACCGCGTTTGTGCTTCTACACCGACA 340
              || |||    ||  || || |  ||| |||| ||   ||    ||  ||  |||| ||||
Db    317 TTGACTGGAGAATCGGAGATCGGCCTTGGTGTAGGATAACTGCGCTAGTGTTTACATCAACA 376

Qy    341 CCTACGGCGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGTGTCT 400
              | |   ||  ||| |||  ||||  || ||  ||| ||| | |  | ||| |||| |
Db    377 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGGCTTCATTGCTGTGTGTGC 436

Qy    401 GTGCCCACTGGGGCCCCGCCTTCGCGCACGGCTGGTCTTGCTGGTCTTGCGTCTGGCCA 460
               | ||  | ||   |   || ||   ||||  |  |  || |||   ||| |
Db    437 ACCCTGTACGCGTGCTACAACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG 496

Qy    461 TCTGGACCTTGGTGTCGTGCAGAGCAGCCCTTGCTTCTGATGACCAAGCCGC 520
              ||||  |   || || | |   |  | |||  || |  |||| | ||| |
Db    497 TCTGGATTCTAGTATTGCTCAGACTTTATGACCTGCCTGAGTATTGACCGGCTCAAAGCAGG 556

Qy    521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTTCAGTGCTGGTGGCCGTGGTGG 580
              || | ||| |||  |||||  |||  ||||| || || || |  | ||||  ||| |  |
Db    557 AGGCTGAAAGGATTACATGCATGGAGTATGTCCATTGGCTTCGTGTGGGCCAGTGGGCCAGTGGTAT 616

Qy    581 TCATGGTCCTCGGTGGTGCCATTGGCCATTGCGTGTGGGACCAGTGGACCATCCTGTCCT 640
              |||||| |  |||| || | ||| |    ||   |||| |||  || | |||| |
Db    617 GGATTCTGCTTGGGGCATGTTTCATAGGATATGAGAGCCAGTGGATAGGAATATGATTCCATCT 676

Qy    641 GCTATATGAAGATCACCTGGAGTCTGCTACGCTGCTGCTCGATGCTGTGTGGCCGTGGTGG 700
              |||||||||| | | |||  | ||| | |  | ||  || |||   | ||| | |
Db    677 GCTATTCTCAGATCGTGCTCAAACTCTTCAGAACTTTGAAGAACATGCAAAACAAAAACCCACTCCA 736

Qy    701 GGAAAGGACGACCACTGGGCAGGTGCCTGCTTACGGTGGCCGTGGTGG 760
              | ||||| ||    | ||    | |  |  |  | ||| || | |||||
Db    737 AATCTGGTGTAAACAAAAAGGCTCTCAACATCAAGAGTTCAACACAATTATTCTTATTATTGTGTTTGTTC 796

Qy    761 TCTGCTTCAGCCCCTACCACCTTACCATGGTCGACGAAAGAAATTCACGTGAAATGACA 820
              ||||||||| | ||| | |||||| || ||  | |||  |||||  | | |||| ||||
Db    797 TCTGTTTCACACTTACCATGTTGCAATTATTCAACATATGATTGATTAAGAAGCTTCGTTTCT 856

Qy    821 TGCCATCC------TGTGCCGAGCGGAGGGCTTTCAGGCACCGTGG 874
              |||||||       |||||| ||   || |  |  | ||||| |||
Db    857 CTAATTTCCTGGAATGTAGCACAAAGACATTCGTTCCTGCACTTTACAGTAT 916

Qy    875 ACAGAAATGGCTCCTGGGCATTTAAGCTCAAAGGTCTTCCTCCTCCT 934
              | | |||| |||  |||  |||  ||||| |||  |||  |||| |||
Db    917 CCCTCATGAACATGAACTGTGGCATTACCCAATCATTTACTTCTTTGCATCCACCCATT 976

Qy    935 ATAAGAAAGGTTATGAGAGATGCTGAAACGGCAAGTGGTATGTATGCCAAGTTCTCAGTGTCTG 994
              |  | ||||   |  ||  ||||  |||| | | | ||    |    || ||||  ||  ||
Db    944 ... 1003

Qy    995 CCTCCTCACCCCAGGAAAAGCTTCTTCAGGAAACACCA 1032
              |||  |||    | ||| || || ||||  ||| ||||
Db   1004 TGAAGTCAGCCCCTGAAGAAAATTCACGTGAAATGACA 1041
```

```
RESULT 9
US-10-305-720-1279
;   Sequence 1279, Application US/10305720
;   Publication No. US20040010136A1
;   GENERAL INFORMATION:
;   APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
;   TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expressio
;   FILE REFERENCE: PA-0002-1 CON
;   CURRENT APPLICATION NUMBER: US/10/305,720
```

```
Qy   935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCCT 994
             ||||| |||||  ||||||  |||| || |||||||| |||||   |||||||||||
Db   977 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1036

Qy   995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
             |||| ||||||
Db  1037 TGAAGTCAGCCCCTGAAGAAATTCACGTGAAATGACA 1074
```

RESULT 10
US-10-247-671-39
; Sequence 39, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247,671
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
;
; SEQ ID NO 39
; LENGTH: 2477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2003019472lA1 057385.565
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2161
; OTHER INFORMATION: a, t, c, g, or other
US-10-247-671-39

Query Match   16.7%; Score 184.8; DB 14; Length 2477;
Best Local Similarity   50.6%; Pred. No. 7.1e-43;
Matches 475; Conservative   0; Mismatches 457; Indels   6; Gaps   1;

```
Qy   101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGAAACATCCTTGCCCTTTGCC 160
             ||| ||| || ||| || ||  |||| ||| |||  ||||| || ||||||||||
Db   548 CTCTGCATTACAGCCCTGCTCTTCATCATTGGGCTGGGAAACTTACTAGCCTTGGTCG 607

Qy   161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGGCATCTACCTGGTGCACCTGGCTG 220
             ||||| ||||| |||| |||| | |||| | |||| ||||| ||  ||| |||| ||
Db   608 TCATTGTTCAAAACAGGAAAAAATCAACTCAACTCTATTCAACAAATTTGGTTA 667

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAGAAGGGTGTCTTATGTGCTGGGCT 280
             ||||||| ||  | |||| ||  ||||| || || |  || || || ||| ||||||
Db   668 TTTCTGATATACTTTTTACCACCGGTTGCCTACGACGAATAGCCTACTATGCAATGGGCT 727

Qy   281 CCAGGTCGGGCTTTCGGCAAGGGGCTCGCAGGCTGCAGGCGTTTGTGCTCTACACCGGACA 340
             |||| || |||  |||  ||  |||||  |||||| ||  || ||  | | ||||  ||
Db   728 TTGACTGGAGAATCGGAGATCGGAGATGCCCTTGTGTGTAGGATAACTGCGCTAGTGTTTTACATCAACA 787

Qy   341 CCTACGGGGGGTTCACCTACGTCATGGGCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
             |||| || ||||  ||||| ||  ||||  ||  || |||||||  ||||   ||||||||
Db   788 CATATGCAGGTGTGAACTTTATGACCGCTTTATGACCGCTTCATTGCTGTGGTGC 847

Qy   401 GTGCCCACTGGGCCCGCGCCTCCGCAGGCTGCGGCGCCAGGCTGCTCGCGTGGGCCA 460
             |||  | |  |||| |||  || || |||||| |  || |||| || |||||||  |
Db   848 ACCCTCTACGTCACAACAGATCACAACAGGATTGAACATGCAAAAAGGCGTGTGCATATTTG 907

Qy   461 TCTGGACCTTGGTGTCTGCTGCAGACGAGCCCTGCTCTTGGCTCTTGATGCCCATGACCAGC 520
             |||| || || |||| |||| || |  | |||| ||   | ||  || || ||||||| |||
Db   908 TCTGGATTCTAGTATTTGCTCAGACACTTCTTCATCAAGTCAAAGCAG 967

Qy   521 TGGTGGCAAGCTGGCCAGCTGGCCCGCATGCAGCAGTACGCAGTGAGTCAGTGCTGCGCCCC 580
             ||  ||| || |||  | |||| ||  ||  | ||||| |||| ||  | |   ||||||||
Db   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGGTGTGTTATGTGCTGGGCT 280
```

```
Db   968 AGGCTGAAAGGATTACATGCAGTCAGTGGAGTATCAAACTTTGAAGAAACTAAAATCTTCCTT 1027
             |||||||| |||  || ||| | ||| |||| || ||||  ||||| || || || |||||
Qy   581 TCATGGTCCTGGTGGCCTTTGCCATTGGCCATTGTGCTCGTGGGCCAGTGGGATCATCGTCT 640

Db  1028 GGATTCTGCTTGGGCATGTTCATAGGATGATGTACTTCACCTATAATCATTCCATCT 1087
             |||||||| || || ||   ||| ||||  || ||||| |  | | |||||| ||||
Qy   641 GCTATATGAGATCACCTGGAAGTCTGTCGAGCACAGCTGGAGCACCCAGTGACCAGCA 700

Db  1088 GCTATTCTCAGATCGTCGTGCAACTCTTCAGAAACTGCCAAACAAAACCCACTCACTGAGA 1147
             ||||| |||| || | ||  |  ||||||||  || ||  ||| |||  || ||| |||
Qy   701 GGAAAGACGCGCACTGGCGAGGCTGCCTGCTTACGCTGCCTGATGCTGGTCGGTGGTGG 760

Db  1148 AATCTGGTGAAACAAAAGGCTCTCAACACCAATTATTCTTATTATTGTTGTGTTTGTTC 1207
             || |||||  | |||| | |||| |  || ||| | || ||||  | || || || ||
Qy   761 TCTGCTTCAGCCCTACCACCTCAACATCAAGAGTCGAGAGGGATGCTCCACC 820

Db  1208 TCTGTTTCACACTTACCATGTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 1267
             ||||| || || ||   || ||  || ||| | ||||| || |||| |||| |||||
Qy   821 TGCCATCC------TGTGCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGCGCCACCGTGG 874

Db  1268 CTAATTTCCTGGAATGTAGCCAAAGACATTCGTTCCAGATTTCTCTGCACTTTACAGTAT 1327
             |||||| | || |||||  | ||||| ||| ||| |||||| | ||||| || |||| ||
Qy   875 CCCTCATGAACATGAACTGTGCATACCCCCAATCATTACTTCTTTGCATCACCCACCATT 934

Db  1328 GCCTGATGAACTTCATTGCTGCATGGACCCTTTTTATCACTTCTTTGCATGTAAAGGGT 1387
             || |||||||| || |  ||  ||||| |||| |||||  |||| | ||| || |||
Qy   935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCT 994

Db  1388 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCTG 1447
             |||||| |||||  ||||||| |||||  | |||||| |||| | ||  |||| |||||
Qy   995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032

Db  1448 TGAAGTCAGCCCCTGAAGAAATTCACGTGAAATGACA 1485
```

RESULT 11
US-10-251-385-205
; Sequence 205, Application US/10251385
; Publication No. US20030105292A1
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
; TITLE OF INVENTION: Protein-Coupled
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/10/251,385
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US/09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 205
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-251-385-205

Query Match   16.6%; Score 183.2; DB 14; Length 1086;
Best Local Similarity   50.5%; Pred. No. 1.4e-42;
Matches 474; Conservative   0; Mismatches 458; Indels   6; Gaps   1;

```
Qy   101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGAAACATCCTTGCCCTTTGCC 160
             ||| ||| || ||| || ||  |||| ||| |||  ||||| || ||||||||||
Db   104 CTCTGCATTACAGCCCTGCTCTTCATCATTGGGCTGGGAAACTTACTAGCCTTGGTCG 163

Qy   161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGGCATCTACCTGGTGCACCTGGCTG 220
             ||||| ||||| |||| |||| | |||| | |||| ||||| ||  ||| |||| ||
Db   164 TCATTGTTCAAAACAGGAAAAAATCAACTCAACTCTATTCAACAAATTTGGTGA 223

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGGTGTGTTATGTGCTGGGCT 280
```

; SEQ ID NO 111
; LENGTH: 1638
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-567A-111

```
Query Match          16.6%;   Score 183.2;   DB 14;   Length 1638;
Best Local Similarity 50.5%;   Pred. No. 1.7e-42;
Matches 474;   Conservative 0;   Mismatches 458;   Indels 6;   Gaps 1;
```

```
Qy  101 CTCTGTTCTACACAGCCTCCTGTCTTCAGTGCCTGCCCTTGCC 160
        |||||| ||| |||||||| ||| ||||| ||| ||
Db  137 CTCTGCATTACAGCCTGCTCTTCATCATTGGGCTTCGTGGGAAACTTACTAGCCTTGGTCG 196

Qy  161 TTACCTGTCAAAAGACAGGAAGATCAACTGCACAGGCATCTACCTGGTCGCACCTGGCTG 220
        || |||||||||| |||||||   |||||| || ||    |||    |||||
Db  197 TCATTGTTCAAAACAGGAAAAAAATCAACTTCTATTCAACAAATTTGGTGA 256

Qy  221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTGTATGTGCTGGGTGC 280
        |||||| ||| |       ||||||   ||||||  ||| || || |
Db  257 TTTCTGATATACTTTTTACCACCGCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 316

Qy  281 CCAGCTGGCCTTCGGCAAGGGGCTGCTGCAGCGTGACGGCGTTTGTGCTCTACACCGACA 340
        ||||||||||||| |||  || || |||| |  ||| ||||| ||| |||||| ||||| |
Db  317 TTGACTGGAGAATCGGAGATGCCCTGTGTGTGTAGGAGGATAAACTGCGCCTAGTGTTTTTACATCAACA 376

Qy  341 CCTACGGGGGGGGTCTCACCTCATGGCCGTGTGTGAGCGTGAGACCATTACCCAGCTGTGGTCT 400
        | ||| ||| ||| |    |||| ||| ||||| |||  |||| || ||| | ||| ||| ||
Db  377 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGTCTGCTGGTGTGC 436

Qy  401 GTGTCCCACTGGGGCCCGGCCTCCGGCGTCCGCCAGCGCACGGCTGGTCTGCTGTCGGCCA 460
        ||| ||||||||||  |  || |  || ||  |||||| |  || || |||| ||||| | |
Db  437 ACCCTTCTACCGTACAACAGATAAAGGATTGAACATGCAAGGGCGTGTGCATATTTG 496

Qy  461 TCTGGACCTTGGTGCTGCTGCTCGACACGATGGCCCTGCTTGATGCCCATGACCAAGCCGC 520
        ||||| |||| |||| ||| ||  |||| |      ||||| |  ||  |||| |||||| |
Db  497 TCTGGATTCTAGTATTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG 556

Qy  521 TGGTGGGCAAGCTGCCCTGGCCTGCGATGAGTACAGCAGCAGCTCAGTGAGTCCTGCCCCC 580
        ||||||||||  || || |||| || ||| ||| |||| || |    || |||| ||  |
Db  557 AGGCTGAAAGGATTACATGCAGCATGAGTATCCAAACTTTGAAGAAAACTAAATCTCTTCCCT 616

Qy  581 TCATGGTGCCTCGTGGCCTTGCCATCCCCATTGCTCCTCGTTCTTTGCTCATCACCCATT 640
        || ||| |||| |||||| || |||| || ||| || |    ||    || |  ||| |
Db  617 GGATTCTGCTTGGGGCATGTTCATAGGATATGTACTTCCACTTATAGGAATGTGCTGCGCTCTCATCT 676

Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGGAGACCCAGTGACCAGCA 700
        ||||| || |||| |  |||||| ||| ||  ||  ||  ||| ||| | || ||||| |||
Db  677 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 736

Qy  701 GGAAAGGACGCCACTGGCGAGGCTGCCTGCTTACGCGTGATGCTGGTGGCCGTGGTGG 760
        ||||| |  |||||| ||| ||||| ||||| | ||||| |||| || ||||| |||
Db  737 AATCTGGTGTAAACAAAAGGCTCTCAACACAATTATTCTTATTATTGTTGGTGGTGTTGTTC 796

Qy  761 TCTGCTTCAGCCCCTCCACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 856
        |||| ||||| |  || |||||| ||||| || ||   ||||||| |||| ||||||||| |||| |
Db  797 TCTGTTTCACACATGCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT ...

Qy  821 TGCCATCC------TGTGCCGAGCGGAGGGCCTTTCTCCCTTCAGCCACCGTGG 874
        ||||||||      ||||| |||| ||| | |  ||  || ||||| || ||||
Db  857 CTAATTTCCTGAAATGTAGCCAAAGACATTCGTTCCTGCACTTTACAGTAT 916

Qy  875 CCCTCATGAACATGAACTGTGCCATTACCCAATCATTACTTCTTTGCATCCACCCATT 934
        |||| |||||| ||||| ||  |||  |  ||  ||||| |||| ||||||||  ||||
Db  917 GCCTGATGAGAACTTCAATTGCTGCATGGACCCTTTATCCACTTCTTTGCATGTAAAGGGT 976

Qy  935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCT 994
        |||||||| ||| ||  |||| || |||| |  || |||||  |||| ||  || ||
Db  977 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTTGTATCGTTCTTCAGTGCTG 1036

Qy  995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
        ||| |  | ||| ||| ||||||    ||| |||||||
Db 1037 TGAAGTCAGCCCCTGAAGAAAATTTCACGTGAAATGACA 1074
```

```
Db  224 TTTCTGATATACTTTTTACCACGGCTTGCCTACACGAATAGCCTACTATGCAATGGGCT 283
Qy  281 CCAGCTGGCCTTCGGCAAGGGGCTGCTGCAGCGGCTGGTCTGCTGTCGCTACACCGACA 340
Db  284 TTGACTGGAGAATCGGAGATGCCCTGTGTGTGTAGGATAAACTGCGCCTAGTTTTTACATCAACA 343
Qy  341 CCTACGGGGGGGTCTACCTCATGGCCGTGTGTGAGCGTGAGACCATTACCCAGCTGTGGTCT 400
Db  344 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGTCTGCTGGTGC 403
Qy  401 GTGTCCCGGGGGCCCGGCCTCCGGCGTCCGCCAGCGCACGGCTGGTCTGCTGGCCA 460
Db  404 ACCCTTCTACGTACAACAAGATAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG 463
Qy  461 TCTGGACCTTGGTGCTGCTGCTCGACACGATGGCCCTTGCTCTTGATGCCCATGACCAAGCCGC 520
Db  464 TCTGGATTCTAGTATTTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG 523
Qy  521 TGGTGGGCAAGCTGGCCTTGGCATGGATACGCAGCAGCAGCATGAGTCAGTGCGCCTGCCCCC 580
Db  524 AGGCTGAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT 583
Qy  581 TCATGGTGCCTCGTGGCCTTCATGGCCATTCCCATTGCTCCTCGTTCTTTGCTCATCACCATTCCATCT 640
Db  584 GGATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 643
Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGTGCAGCACAGTCGGGAGGACCAGTGACCAGCA 700
Db  644 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 703
Qy  701 GGAAAGGACGCCACTGGCGAGGCTGCCTGCTTACGCGTGATGCTGGTGGCGTGGTGG 760
Db  704 AATCTGGTGTAAACAAAAGGCTAAACACAATTATTCTTATTATTGTTGGTGGTGTTGTTC 763
Qy  761 TCTGCTTCAGCCCCCTACCGCCTACCTTCAAGCAGTTCATGGCGAGAGGGATGCTTCCACC 820
Db  764 TCTGTTTCACACATGCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT 823
Qy  821 TGCCATCC------TGTGCCGAGCGGAGGGCCTTTCTCCCTTCAGCCACCGTGG 874
Db  824 CTAATTTCCTGGAAATGTAGCCAAAGACATTCGTTCCTGTTCTTCAGATTTCAGTAT 883
Qy  875 CCCTCATGAACATGAACTGGCATTACCCCAATCATTACTTCTTTGCATCCACCCATT 934
Db  884 GCCTGATGAGAACTTCAATTGCTGCATGGACCCTTTATCCACTTCTTTGCATGTAAAGGGT 943
Qy  935 ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCCT 994
Db  944 ATAAGAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTTGTATCGTTCTTCAGTGCTG 1003
Qy  995 CCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
Db 1004 TGAAGTCAGCCCCTGAAGAAAATTTCACGTGAAATGACA 1041
```

RESULT 13
US-10-350-923B-29
; Sequence 29, Application US/10350923B
; Publication No. US20040002084A1
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Ma, Xiao Jun
; TITLE OF INVENTION: Nucleic acid, polypeptides, vectors, and cells derived from activ
; TITLE OF INVENTION: eosinophil cells
; FILE REFERENCE: S03157-01
; CURRENT APPLICATION NUMBER: US/10/350,923B
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US/ 09/454,280
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: PCT/US99/28773
; PRIOR FILING DATE: 1999-06-12
; PRIOR APPLICATION NUMBER: US 60/111,006
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 1556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-350-923B-29

Query Match   15.6%;   Score 172.2; DB 15; Length 1556;
Best Local Similarity 50.5%; Pred. No. 2.6e-39;
Matches 474; Conservative 0; Mismatches 458; Indels 7; Gaps 2;

```
Qy   101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
             |||||  | || |||||| || ||||| | |  | || |||||| |  || |  ||
Db   234 CTCTGCATTACAGCCTCGTCTTCATCATTGGGCTCGTTGGGAAACTTACTAGCCCTTGGTCG 293

Qy   161 TTACCTGTCAAAAGACAGGAGAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
             | | || |||||| ||| | || ||||      | |   ||   |
Db   294 TCATTGTTCAAAACAGGAAAAAAATCAACTCTACCACCCTCTATTCAACAAATTGGTGA 353

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTT-ACCGGGAAGGGTGGTGTTGTTATGTGCTGGGC 279
           | ||||  | |  ||||| ||    | | |  |  ||| | |  |   ||  |||||
Db   354 TTTCTGATATACTTTTTACCACCGCGTTTGGCCTACACGAATAGCCTACTATGCAATGGGC 413

Qy   280 TCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAAGCCTGACGCGGCGTTTGCTCCTACACCGAC 339
           | | ||| | |  || | |  |  ||  |  ||| | | |  | | ||  |  ||| |||
Db   414 TTTGACTGGAGAATCGGAGATGCCTTGTGTAGGATAACTGCGGCTAGTGTTTTACATCAAC 473

Qy   340 ACCTACGGGGGGGTCTACCTCCATGGCCGTGTGAGCGTGGACCATTACCCAGCTGGGTC 399
           || | |  | |||  || | |  | ||   |    || | |||    |  ||| |
Db   474 ACATATGCAGGTGTGAACTTTATGACCTGCCGAGTATTGACCGCTTCATTGCTCGTGGTG 533

Qy   400 TGTGCCCACTGGGGCCCGGCCCTCCGGACAGCGTGGCTGGGCTGGTCGCGTGGGCC 459
            | || |||  | |     | |  |  || | || | | | |  |   | || ||
Db   534 CACCCTCACGCTACAGCTACAAGATAAAAGGATTGAACATGCAAAAGGCGTGCATATTT 593

Qy   460 ATCTGGACCTTGGTGCTGCTGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCAG 519
           | |||| | | || | | |  | || |  | |  | |  |||| |  | |  ||||||
Db   594 GTCTGGATTCTAGTATTTGCTCAGACACACTCCCACTCCTCATCACCCTATGTCAAAGCAG 653

Qy   520 CTGGTGGGCAAGTGGCCTGCATGGACTACAGCAGCATGGAGTCATGTCCTCGGGCTGCCC 579
            | |  | | |  ||| | | ||  | | |  || ||||||||||  |  | | ||  ||
Db   654 GAGGCTGAAAGGATTACATCATGGAGTCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCC 713

Qy   580 CTCATGGTCCTGGTGGGCCTTGCGCCTTCGTGTGGGGGATCATCCTGTCC 639
           | |||| |  | || ||  |  | | ||  |  | |||||| | |  | |
Db   714 TGGATTCTGCTTGGGCATGTTTCATTGGGGGCATGTTCACTTATAAATGTCATTCTCATC 773

Qy   640 TGCTATATGAAGATCACCTGGAAGCTGGAAGCTGTGCAGCAGACCAGTGACCAGC 699
           |||||   ||| | |  | |  |  | | | |  || | |  ||| | ||| |||
Db   774 TGCTATTCTCAGATCTGCTGCAAACTCTTCAGAATCTTCAGAACAAACAAACCCACTCACTGAG 833

Qy   700 AGGAAAGGACGGACGCCACTGGCCAGCCGCTGCTTACGCTGCTTACGCTGCTGGTGGCCGTGGTG 759
             | |||     | ||| |||| |  || | |   | ||  || | ||   | ||| || |||||
Db   834 AAATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGTTGTGGTTTTGTT 893

Qy   760 GTCTGCTTCAGCCCTCCACCTCAACATCAAGCAGTCAGTCATGGCGGAGAGGATGCTCCAC 819
            |||| |||   | |  || | |   | | |  |  |  | ||| |  ||| |  |  |||
Db   894 CTCTGTTTCACACCTTACCACTGTGCAATTATTCAACATATGCATATGATTAAGAAGCTTCGTTTC 953

Qy   820 CTGGCCATCC-----TGTGCCGAGCGAGGGCTTCTTACTGTCCCTTCAGGCCACCGTG 873
            |||||||||     ||||||||||||| ||| |||| || |  ||| |||||||| |
Db   954 TCTAATTTCCTGGAATGTAGCCAAAGACATTCGGATTTCTGCGACATTTTACAGTA 1013

Qy   874 GCCCTCATGAACATGAACTGTGCGATTACCCCAATCATTACTTCTTGCATCCACCCAT 933
           |||||||||||| ||||  | || | |||   |||||| | | |  || | |||| ||
Db  1014 TGCCTGATGAACTTCATTGCTGCATGGACCCTTTTATCTACTTCTTTGCATGTAAAGGG 1073

Qy   934 TACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAGGGTCTTCCTCCTCCTCCTCC 993
           ||||  |||  | |  || | ||| |||| ||||||   |    | | |   || ||
Db  1074 TATAAGAAAGGTTATGAGGATGCTGAAACGGCAAGTCAGTGTATCGATTTCTAGTGCT 1133

Qy   994 TCCTCCTCCACCCCAGGAAAAGCTTCTTCAGAAACACCA 1032
             | |  | |||||  |||||| || || ||| | ||  |
Db  1134 GTGAAGTCAGCCCTGAAGAAAATTCACGTGAAATGACA 1172
```

RESULT 14
US-09-812-102-90/c
; Sequence 90, Application US/09812102
; Patent No. US20020055179A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E
; TITLE OF INVENTION: No. US20020055179A1e1 G-Protein Coupled Receptor Homologs
; FILE REFERENCE: 5800-41 03S800/183478
; CURRENT APPLICATION NUMBER: US/09/812,102
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/364,769
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: GPCR-RHODOPSIN
US-09-812-102-90

Query Match   13.4%;   Score 148; DB 9; Length 594;
Best Local Similarity 96.4%; Pred. No. 1.8e-32;
Matches 162; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
Qy   229 CTGCTGTTCACCGTGGCCTTACCGTGAGGGTGGTGTTATGTGCTGGGCTCCAGCTGG 288
           ||||||||| |||||||||||||  |||||| |||||| ||||||||||||||||||
Db   590 CTGCTGTTCCACCTGGCCTTACCGGAAAGGGTGTGTTATGTGCTGGGCTCCAGCTGG 531

Qy   289 CCTTTCGGCAAGGGGCTCTGCGAGGCTGACGCGGTTTGTGCTCTACACCGACACCTACGGG 348
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   530 -CTTTCGGCAAGGGGCTCTGCCAGGCTGACGCGGTTTGTGCTCTACACCGACACCTACGGG 472

Qy   349 GGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 396
           |||||||||||||||||||||||||||||||||||||||||||| |||
Db   471 GGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCCGTG 424
```

RESULT 15
US-09-918-995-10498
; Sequence 10498, Application US/0991899S
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995

```
;  CURRENT FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: US/09/235,076
;  PRIOR FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 38054
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 10498
;  LENGTH: 548
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(548)
;  OTHER INFORMATION: n = A,T,C or G
US-09-918-995-10498
```

```
Query Match       11.2%;   Score 123.4;  DB 10;  Length 548;
Best Local Similarity  52.8%;   Pred. No. 2.5e-25;
Matches  262;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy    167 GTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTGTGTCTG 226
              |||||||| ||||||||| || |  || ||||| ||||   |||  |||||||| |||
Db     53 GNTANAACAGGAGAGAAATCAACTCTACCACCCTCTATTCAACAAATTTGGGTATTTCTG 112

Qy    227 ACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTGTTATGTGCTGGGCTCCAGCT 286
              ||||| || ||||| ||||| |||| ||| | | ||  ||||| ||| |||||||| ||
Db    113 ATATACTTTTTACCACCGCTTTGCCTACACGAATAGCCTACATGCAATGGGCTTTGACT 172

Qy    287 GGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTGTGCTCTCACCGACACCTACG 346
              || |||  ||| |||| ||||  ||| || || ||||| ||||   ||||||||||||
Db    173 GGAGAATCGGACGACGCCTTGCGTAGGATAACTGCGCTAGTGTTTTACATCAACACATATG 232

Qy    347 GGGGGGTCTCACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCC 406
              |||     |||  | |   |||| ||||| | ||  |||||||  ||||||||| ||
Db    233 CAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGCTGTGGTGCACCCTC 292

Qy    407 ACTGGGGCCCGCCGCCTCCGCACGGCTGGCCGCCGGCCAGGCTGGTCTGCGTGGCCATCTGGA 466
              |||  ||  |   |  | ||| || | |||   |||    |||||| | | ||||||||||
Db    293 TACGCTACACAAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTGTCTGGA 352

Qy    467 CCTTGGTGCTGCGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGG 526
              |||  ||  || ||| || || || ||||| ||  || | |  ||||||||| |||||
Db    353 TTCTAGTATTTGCTCAGACACTCCCACTCCCATCATCAACCCCTATGTCAAAGCAGGAGGCTG 412

Qy    527 GCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCCAGTCCTGCGGGCTGCCCCCTCATGG 586
              || ||||| ||||    ||| |   || ||||   || |||||| |||| |||
Db    413 AAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTTTCCCTGGATTC 472

Qy    587 TCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGGATCATCCTGTCCTGCTATA 646
              |||| || ||||| ||     |||||| || |||  |||||||| ||||| ||||| ||||
Db    473 TGCCTGNGGCATGTTTCATAGGATATGTGACTTCCCACTTATAATCATTCTCATCTGCTATT 532

Qy    647 TGAAGATCACCTGGAA 662
              ||||| || ||| |||
Db    533 CTCAGATCTGCTGCAA 548
```

```
Search completed: February 21, 2004, 21:36:27
Job time : 474 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 19:11:15 ; Search time 2793 Seconds
                                (without alignments)
                                11803.741 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaaactaggcccta........grgtcattgctttatgact 1104

Scoring table: IDENTITY.NUC
                Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
         1: em_estba:*
         2: em_esthum:*
         3: em_estin:*
         4: em_estmu:*
         5: em_estov:*
         6: em_estpl:*
         7: em_estro:*
         8: em_htc:*
         9: gb_est1:*
        10: gb_est2:*
        11: gb_htc:*
        12: gb_est3:*
        13: gb_est4:*
        14: gb_est5:*
        15: em_estfun:*
        16: em_estom:*
        17: em_gss_hum:*
        18: em_gss_inv:*
        19: em_gss_pln:*
        20: em_gss_vrt:*
        21: em_gss_fun:*
        22: em_gss_mam:*
        23: em_gss_mus:*
        24: em_gss_pro:*
        25: em_gss_rod:*
        26: em_gss_phg:*
        27: em_gss_vrl:*
        28: gb_gss1:*
        29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	211.2	19.1	1496	11	AK087951	AK087951 Mus muscu
2	180.6	16.4	631	14	CB499474	CB499474 ssalkhc01
3	169.6	15.4	1097	9	AL547762	AL547762 AL547762
4	168.4	15.3	1171	13	BX356015	BX356015 BX356015

5	160.2	14.5	979	9	AL532537	AL532537 AL532537
6	160.2	14.5	1201	13	BX417963	BX417963 BX417963
7	159.4	14.4	636	14	CF359817	CF359817 821090 MA
8	159.4	14.4	1201	8	AL551903	AL551903 AL551903
9	155.6	14.1	1088	12	BM549799	BM549799 AGENCOURT
10	155.2	14.1	715	13	BY765663	BY765663 BY765663
11	151.2	13.7	963	13	BY745873	BY745873 BY745873
12	150.6	13.6	1201	13	BX427288	BX427288 BX427288
13	148	13.4	594	9	AI659965	AI659965 we64e01.x
14	143.4	13.0	662	13	BY729793	BY729793 BY729793
15	140.4	12.7	861	12	BI768868	BI768868 603057774
16	139.4	12.6	660	13	BQ562908	BQ562908 H4078F07-
17	139	12.6	679	13	BY751377	BY751377 BY751377
18	138.4	12.5	611	14	CF913347	CF913347 A0647A02-
19	138.4	12.5	682	13	BY750933	BY750933 BY750933
20	137.2	12.4	933	13	BX453641	BX453641 BX453641
21	133.4	12.1	778	14	CD468929	CD468929 LeukoS3_8
22	133	12.0	614	14	BB637586	BB637586 BB637586
23	124	11.2	678	14	CD687357	CD687357 EST3878 h
24	124	11.2	695	14	CD686410	CD686410 EST2931 h
25	124	11.2	749	12	BI821236	BI821236 603034950
26	122.4	11.1	686	14	CD708189	CD708189 EST24716
27	121.6	11.0	1201	9	AL575476	AL575476 AL575476
28	121.6	11.0	671	14	CD640122	CD640122 AGENCOURT
29	119.6	10.8	682	13	BY750497	BY750497 BY750497
30	119	10.8	620	14	CF181257	CF181257 818066 MA
31	118.4	10.7	628	10	BB636985	BB636985 BB636985
32	118	10.7	574	12	BI009654	BI009654 MR4-RT004
33	117.6	10.7	673	14	CD465254	CD465254 LeukoN1_2
34	115.6	10.5	643	14	CF249695	CF249695 esa003_a0
35	115.4	10.5	673	12	BG547923	BG547923 602576027
36	115	10.4	631	14	CD470715	CD470715 LeukoS4_6
37	113.6	10.3	879	12	BG536922	BG536922 602564666
38	112.4	10.2	660	12	BI394962	BI394962 pgp1n.pk0
39	111.4	10.1	539	14	CA575605	CA575605 K0633CC04-
40	110.6	10.0	532	14	CA576608	CA576608 K0647D04-
41	108.2	9.8	789	14	CD521415	CD521415 AGENCOURT
42	107.6	9.7	679	10	BB663009	BB663009 BB663009
43	106.4	9.6	1019	12	BI517811	BI517811 603042065
44	105.8	9.6	4022	11	AK043873	AK043873 Mus muscu
45	102.6	9.3	700	12	BG483806	BG483806 602503457

ALIGNMENTS

RESULT 1
AK087951
LOCUS   AK087951   1496 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
            full-length enriched library, clone:E430001G12 product:EBV-INDUCED
            G PROTEIN-COUPLED RECEPTOR 2 (EBI2) homolog [Homo sapiens], full
            insert sequence.
ACCESSION   AK087951
VERSION   AK087951.1 GI:26104619
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE   20499374

```
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,H., Ikegami,T., Kashiwagi,A.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6   (bases 1 to 1496)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
    TITLE   Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp/,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
  COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..1496
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="FANTOM_DB:E430001G12"
                     /db_xref="MGI:2426848"
                     /db_xref="taxon:10090"
                     /clone="E430001G12"
                     /cell_type="thymic cells"
                     /tissue_type="thymus"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days neonate"
     misc_feature    1..1496
                     /note="EBV-INDUCED G PROTEIN-COUPLED RECEPTOR 2 (EBI2)
```

ORIGIN

```
                homolog [Homo sapiens] (SWISSPROT|P32249, evidence: FASTY,
                87.3%ID, 100%length, match=1082)"

Query Match
Best Local Similarity   19.1%;   Score 211.2;   DB 11;   Length 1496;
                        53.1%;   Pred. No. 1.2e-23;
Matches 498;   Conservative 0;   Mismatches 433;   Indels 7;   Gaps 2;

Qy   101  CTCGTTCTACACAGCCCTCCGTGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC  160
               ||||    |||||    ||||||||  ||  ||| ||    |||| ||  ||   ||
Db   170  CTCTGCATTACAGCCTGGTCTTCATCATTGGGCTGGTGTGGGAAACCTGGCCTTGGTTG  229

Qy   161  TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATCTACCTGGTGCACTGGCTG  220
               |||||||| ||   |||| | |||   |   |   | |     |||||||
Db   230  TCATTGTTCAAAACAGAAAAAATCACTCAACCTCTCTATTCAATGAACTTCGTCA  289

Qy   221  TGTCTGACCTGCGTTCCACCGTGGCCTTACCGGGAGGGGGTGGTGTGTTATGTGTGGGCT  280
               ||||||| | |   ||||||  |  |||||  | ||||      |  |||||| |||
Db   290  TTTCTGACATCCTGTTTACCACCAGCTTTACCCACTCGGATAGCCTACTATCGCTGGGGCT  349

Qy   281  CCAGCTGCCCTTTCGGCAAGGGGCTCTGCAGGCTGACGCGGGTTTGTGCTTCTACACCGACA  340
                 |   |  |  ||||  |   |  || |  ||||| | ||| | || | |||  |  |
Db   350  TTGATTGGAGGATAGGTGATGCCCTGTGCCGGGTAACTGCTCGGTGTTCATCATCAACA  409

Qy   341  CCTACGGGGGGGTCTCACCTCATGGCCTGTGTGGAGCGTGGACCATTACCCAGTGTGGTCT  400
               ||||| |||| ||| | ||||||| |||  | |||| ||  || | ||| | ||||| |
Db   410  CGTACGCAGGTGTGAACTTCATGACTTGCTTGAGCATAGACCGCTTCTTCGCTGTGGT-G  468

Qy   401  GTGCCCACTGGGGCCCCGGCCTTCCGCACGGCCCAGGCTGGTCTGGCGTGGCCA  460
               ||||| ||| | ||  ||    |||||||  ||||| ||||| |||| | ||
Db   469  CACCTCTGCGTCACAACAAGATTAAAAGAATCGAATACGCAAAGGGGTGTCCCGTGTCCG  528

Qy   461  TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTTCTTGATGGCCATGACCAAGCCGGC  520
               |||| | |||||| ||| ||| ||   ||||   ||| |     || |||  |||| ||
Db   529  TCTGGATTCTGGTTCTTTGCTCAAAACACTGCGGCTGCTCCCACCCTATGTCTAAGGAGG  588

Qy   521  TGGTGGGCAAGCTGGCCTGCATGGAGTACAGACAGCATGGAGTCAGTCCTCCTCGGGCTGCCCC  580
               ||| || |||| |||  |||||||||||  | ||| |||||||  |    || |||| | ||
Db   589  AGGGAGACAAGACCCACTTGCATGGAGTGATCAAACTTTGAAGGGACAGCAGCGTCCCGCCGT  648

Qy   581  TCATGGTCCTGGCCTTTGCCATTGGCTTCGTGTGGGCCAGTGGGATCATCGTGTCTCCT  640
               ||| |  || | |  ||||| || |||| |||| | |||||| || || | |||| |||
Db   649  GGATTCTGCTCGGAGCCTGTCTGCGGGTCACGTGCTGCCTATCACAGTCATTCTCCTGT  708

Qy   641  GCTATATGAAGATCACCTGGAAGCTGTGCAGACAGCAGTCGGGAGGACCAGTGACCAGCA  700
               | |||| ||| |||| ||| || |||||||| ||||| ||||| ||||||||||||| |
Db   709  GTTACTCTCAGATCTGCTGCAAACTCTTCAGGACTGCCAAGCAGAACCCACTCACCGAGA  768

Qy   701  GGAAAAGGACGCCACTGGCGAGGCTGCCTGCTTACCGTGCTCTGATGCTGGTGGCCGTGGTGG  760
               || | |||| || |   ||| ||| | |  ||| ||||| |  | | | ||| || ||
Db   769  AATCTGGTGAACAAAAAGGCTCTCAACACAATTATCCTCACAATTGTCGTGTTCATCCC  828

Qy   761  TCTGCTTCAGCCCCTACCACCTCCACAACTCAGCAGTCATGGCGAGGAGGATGCTCCACC  820
               |||||||| || || |||||| | | ||| ||||| || || || |||||||||| |||
Db   829  TGTGCTTCACGCCTACCACGTGGCATCATTCAGCACATGATAAAGATGCTCTGCTCCCC  888

Qy   821  TGCCATCC------TGTTGCCAGCGGCGAGGGGCTTTCTTACTGTCCCTCAGGCACCGTGG  874
               ||  |          ||| || | || |||  ||||||| |||  ||||||| | |||| |
Db   889  CTGAGCCCTGGAGTGTGGGCGGAGCATTCTTCCAGATCCTCTGCACTTCACGGTGT  948

Qy   875  CCCTCATGAACATGAACTGTGGCATTACCCCAATCATTACTTGCTTGCATCCACCATT  934
               |||| ||||||| |||| | | || ||   ||| |||| | |  ||||  |  || ||
Db   949  GCCTGATGAACTTCAACTTCGATGACCCGTTCATATACTTCTTTGCATGCAAAGGGT  1008

Qy   935  ACAGGAAATGGCTCCTGGGCATTTAAAGCTCAAAAGGGTCTTCAAAGGGTCTCCTCCTCCT  994
               |  |||||  |  || | ||||||||| | |||||| ||| ||||| | | |   ||| |
Db  1009  ATARAGAAAGGTCATGAAGATGCTCAAAACCTCAAGTGAGTGTGTCGATTCCCAGCGCAG  1068

Qy   995  CCTCCTCCACCCCAGGAAAAGTTCTTCAGAAACACCA  1032
               | | || |||||||||| |||| ||||||  |  |
Db  1069  TGAGGTCAGCCCCTGAAGAGAGAATTCGCGGGAAATGACA  1106
```

```
RESULT 2
CB499474
LOCUS       CB499474     631 bp     mRNA     linear     EST 16-MAY-2003
DEFINITION  ssalkhc013018 kidney Salmo salar cDNA, mRNA sequence.
ACCESSION   CB499474
VERSION     CB499474.1  GI:29310700
KEYWORDS    EST.
SOURCE      Salmo salar (Atlantic salmon)
ORGANISM    Salmo salar
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
REFERENCE   1 (bases 1 to 631)
AUTHORS     GRASP Consortium, Davidson,W.S., Koop,B.F. and
            http://web.uvic.ca/cbr/grasp.
TITLE       A survey of Salmo salar transcripts from high complexity cDNA
            libraries
JOURNAL     Unpublished (2002)
COMMENT     Contact: Koop BF
            Centre for Biomedical Research
            University of Victoria
            PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
            Tel: 250 472 4067
            Fax: 250 472 4075
            Email: bkoop@uvic.ca
            Centre for Biomedical Research, University of Victoria cDNA
            preparation and sequencing: Roberto Alberto, Marianne
            Beetz-Sargent, Maura Busby, Peter Hunt, Linda McKinnel, BF Koop.
            bioinformatics: Gordon D Brown.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Salmo salar"
                     /mol_type="mRNA"
                     /strain="McConnell"
                     /db_xref="taxon:8030"
                     /clone_lib="kidney"
                     /note="Vector: pBlueScriptIISK+; Library Creator: Matthew
                     L Rise ; Atlantic salmon tissue contributors: Carlo Biagi,
                     Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
                     Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
                     B.C.), Rachel Roper (University of Victoria)"
ORIGIN

Query Match      16.4%;  Score 180.6; DB 14; Length 631;
Best Local Similarity  56.6%; Pred. No. 2.5e-27;
Matches  355; Conservative  0; Mismatches 269; Indels  3; Gaps  1;

Qy  178 AGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCGTGTGTCTGACCTGCTGTTC 237
            || |||||| |||  || ||  ||||| |  |||   |||||| | ||  ||| | |||||
Db    3 AAGAAGATTAACTCCACCACGCTCTACTCTGCCAACCTGGTCATCTCCGACATCCTGTTC 62

Qy  238 ACCGTGGCCTTACCGGGAGAGGGTGGTGGTGTTATGTGCGTTCCAGCTGGCCTTTCGGC 297
            | |||  | | |||   | | ||||| |  | |  ||||  |||  || ||||| |||||
Db   63 ACACTGTCACTACCGTTACGGATAATCTACTACTACGCCCCTGGGCTTCCACTGGCCCTGGGG 122

Qy  298 AAGGGGCTCTGCAGGCTGACGCGGCGTTGTGCTCTACACCGACACCTACGGGGGGGTCTAC 357
            ||  | ||| |  || || |||| || |  | |||||| ||||  ||||  |||| ||||
Db  123 GAGACCCTGTGTAAGATCGTGGGCTGCTGATCTTCTACATCAATACTTACGCCCGGGGTCAAC 182

Qy  358 CTCATGGCCTGTGTGAGCGTGGACCATTACCCCAGCTGTCGTGCCACTGGGGCCCG 417
            ||||| ||||| ||||| ||||| | ||| || |||  ||||||||||| |||| |
Db  183 TTCATGACCTGCCTCAGTGTGGACCGGTTCATTGCTGCTGCGCGTGGCCATCTGGACCTTGGTGCTG 242

Qy  418 CGCCTCCGCACGGCTGGCCCGCCAGGCGTCTGCGTGGCCATCTGGACCTTGGTGCTG 477
            ||| | ||  |||  |||  ||||    ||  | |||| || ||| |||||| |||||
Db  243 CGCTTCCGTAAGGTCTCCAATGTCCGGTACATTTGCGTCGGGGGTGTGTTGCTGGTCTTA 302

Qy  478 CTGCAGACGATGCCCTGCTCTTGATGCCCATGAC---CAAGCCGCTGGTGGGCAAGCTG 534
            ||| || | |||| || || || ||  || |||    ||||| |||| ||||| | |||
Db  303 ATGCAAACCTGCCCCTCCTGTCCTCCATGCCATGACGAGGAACCTGATGCATGGCTTCATC 362

Qy  535 GCCTGCATGCATGCAGCAGCAGCCATGCAGTCAGTCCTGGGTGCTG 594
            ||||| |||   || | |  |||| |  |||  |  |  |||  || ||
Db  363 TCATTGTTCAAAACAGGAAAAATCAACTCTACCACCCCTATTCAACAATTTGGTGA 341
```

```
Db  363 ACCTGTATGGAGTATCCAACTTCGAACCGGTGCCTAACATCCTCTATATCCTGATTGGC 422
            ||||| ||||   | |  || || || ||  |||  |||||||| || |||||||||||
Qy  595 GCCTTGCCATTGGTTCTGTGGGCAGTGGGGATCATCCTGTCCTGCTATATGAAGATC 654

Db  423 GCCGTATTCCTAGGCTACGGAGTCCCCGTGGTGACCATCCTGGTGTGCTACTCCATACTG 482
            |||||  || | ||||  || ||   || |||||| |  ||| |||| |||| ||||||
Qy  655 ACCTGGAAGCTGTGCAGCACAGCTCGGGAGGAGACCCAGTGACCAGCAGGAAAGACGCCAC 714

Db  483 TGCTGTAAACTCCGCTCCGCGCCAAGGCCAATCAGCTGACGTGGTGGCCGCGGTGGTCTGCTTCACCCCC 774
            ||||   |  |  || || || || || ||| ||| | || ||||| |  | ||||| | ||||   || |
Qy  715 TGGGCAGGCTGCCTGCTTACGCTCGATGGCTGGCGGCCGTGGTGGCCGCGGGTGTTGCTTGCTTCACCCCC 774

Db  543 CAAAAGCCATCGGGGTGATCTGTCTGCG|CTCCCGTGGTGTTTGTGGTCTGTTTGTTCAGCCC 602
            ||||||||||| |||||| ||||||| | | |||||||||||||| || | |||||||||  ||
Qy  543 CAAAAGCCATCGGGGTGATCTGTCTGCG CTCCCGTGGTGTTTGTGGTCTGTTTGTTCAGCCC 602

Qy  775 TACCACCTCAACATCAAGCAGTTCATG 801
            ||| | || |||| ||| |||  |||
Db  603 TATCACATCGACCTCCTCCAATACATG 629
```

```
RESULT 3
AL547762
LOCUS       AL547762     1097 bp     mRNA     linear     EST 31-MAY-2003
DEFINITION  AL547762 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI017YN05 5-PRIME, mRNA sequence.
ACCESSION   AL547762
VERSION     AL547762.2  GI:31269591
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1097)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12882129.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6994.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI017CG03QP1&cluster=6994.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI017CG03QP1.
FEATURES             Location/Qualifiers
     source          1..1097
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI017YN05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match      15.4%;  Score 169.6; DB 9; Length 1097;
Best Local Similarity  49.7%; Pred. No. 5.9e-25;
Matches  434; Conservative 10; Mismatches 424; Indels  6; Gaps  1;

Qy  101 CTCTGTTCTACACAGCCCTCCTGGTTCCTCAGTGCCTCGGGGAAACATCCTTGCCCTTTGCC 160
            ||||| |||    || || |  |||  | |  ||||  ||||||||| ||||  ||| |||
Db  222 CTCTGCATTACAGCCCTGTCTTCATCATTGGCGTCTCGTGGGAAACTTACTAGCCTTTCATC 281

Qy  161 TTACCTGTCAAAAGAGCAGGAGAGATCAACTCAGCACAGGCATCTACCTGGTGCACCTGCTG 220
            || || || |||| || || || ||| || |  |||||| ||| ||||  |  ||  ||| |
Db  282 TCATTGTTCAAAACAGGAAAAATCAACTCTACCACCCCTATTCAACAATTTGGTGA 341
```

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6994.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI007AE01QP2&cluster=6994.r. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI007AE01QP2.

FEATURES   Location/Qualifiers
     source   1..1171
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="CS0DI007YI01"
                   /tissue_type="PLACENTA COT 25-NORMALIZED"
                   /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                   /note="1st strand cDNA was primed with a NotI-oligo(dT)
                   primer, Five prime end enriched, double-strand cDNA was
                   digested with Not I and cloned into the Not I and EcoR V
                   sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match   15.3%;   Score 168.4;   DB 13;   Length 1171;
Best Local Similarity 50.5%;   Pred. No. 1.1e-24;
Matches 420; Conservative   7; Mismatches 398;   Indels   7; Gaps   1;

Qy 101 CTCTGTTCACACAGCCCTCCTGGTCTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
            ||||| |||||| |||| |||||| |||||| ||| | | ||||||| ||||
Db 187 CTCTCGCATTACAGCCTCGTGTTCATCATTGGGCTCGTCGTGGGAAACTTACTAGCCTTGGTCG 246

Qy 161 TTACCTGTCAAAAGACAGGAAGATCAACTGCACAGCAGGATCACCTCACCTTATTCAACAAATTGGTGA 220
           |||||| | | |||||| | || |||||||| |
Db 247 TCATTGTTCAAAACAGGAAAAAATCAACTCTACACCCTCTATTCAACAAATTGGTGA 306

Qy 221 TGTCTGACCTGCTGTTCACCGTGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT 280
            ||||| |||||| |||||||||| ||||||
Db 307 TTTCTGATATACTTTTACCACCGCGTTTGCCTCACACCGCGTTTCATCACCGGACA 366

Qy 281 CCAGCTGGCCTTTCCGCAAGGGCTGCAAGGGCTCGACGGCGTTTGTGCTCTACACCGACA 340
              |||| |||||||||| |||||||| |||||| ||||||||||||| |
Db 367 TTGACTGGAGAATCGGAGATGCCTTGTGTAGGATAACTGCGCTAGTGTTTTTCATCAACA 426

Qy 341 CCTACGGGGGGCTCACCTCAGGCATGGGCTGATGAGCGTGGAGCGTGGACCATTACCCAGCTGTGGTCT 400
         |||| ||| |||| ||||||||||| ||| | | |||| |||||
Db 427 CATATGCAGGTGTGAACTTTATGACCTGGCGTGAGTATTGACCGGTATTGGTGTCGTGGTGC 486

Qy 401 GTGCCCACTGGGGCCGCCTCCGCACGGTCTGGCCGGCGTGGTGGGCCA 460
           ||||||||||||||||| ||||||||| |
Db 487 ACCCTCTACGCTACAACAAGATAAAAGAGATTGAACATGCAAAAGCGTGTGCATATTTG 546

Qy 461 TCTGGACCTTGGTGTGCTGCAGACGATGCCCTGCTTGATGCCCATGACCAAGCCGC 520
          |||||||||| ||| || | ||| | ||||||||||| | | |||| ||| ||
Db 547 TCTGGATTCTTAGTATTTGCTCAGACACTCCTCATCATCCTCATCATCATGTCAAAGCAGG 606

Qy 521 TGGTGGGCAAGCTGGCCTGCTGAGTAGCAGCATGGAGTCAGTGCTCCTCGGGGGCTGCCGC 580
          ||||||||||||| || ||| |
Db 607 AGGCTGAAAGGATTACATGCCATGGCAGTATCCAAACTTTGAAGAAACTAAAATCTTCCCT 666

Qy 581 TCATGGTCTGCTGGTGGCCATTGGCTTCTGTGTGGGCCAGTGGGGATCATCCGTGTCCT 640
          ||||| ||||| ||||| |||||||| ||| ||||||| ||||||| |||||||
Db 667 GGATTCTGCTTGGGCATGTTTCATAGGATATGACTTCCACTTATAATCATTCTCATCT 726

Qy 641 GCTATATGAAGATCAGTGTGCAGCACAGCTCGGGAGGACCCAGTGACCAGCA 700
         ||||||||||||||| | | | || | |||| ||| || |
Db 727 GCTATTCTCAGATCTCTGCAAACTCTTCAGACTGCCAAACAAACCCACTCACTGAGA 786

Qy 701 GGAAAGGACCGCACTGGCAGGCTGCCTTACGCTACGGTCGTGATGCTGGCCGTGTGG 760
           ||||| | | || ||| | | ||| | |||| |
Db 787 AATCTGGTGTAAACAAAAGGCTCTCAACACAATTATTCTTATTATTGTTGTTTGTTC 846

Qy 761 TCTGCTTCAGCCCTCCTACCACCTCAACATCAAGCAGGATGCATGGCGAGAGGGGATGCTCCACC 820
          || ||| || |||||||| |||| || ||

```
Db  435  TTGACTGGAATCGGACGATGCCTTGTGTGTAGGATAACTGCGCTAGTGTTTTACATCAACA  494
              ||||                                              ||||  ||||
Qy  341  CCTACGGGGGGTCTACCTCATGGCCTGGTCTTGTGAGCGTGGACCATTACCCAGCTGTGGTCT  400
              ||||||                ||||         ||||    ||||||
Db  495  CATATGCAGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGCTGCTGGGTGC  554
              ||||            ||||        ||||||
Qy  401  GTGCCACTGGGGCCCGGGCCTCCGGACACGGCAGGCTGGTCTGGTGGCCA  460

Db  555  ACCCTCTACGCTACAACAGATAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG  614
              ||||
Qy  461  TCTGACCTTGGTGTCTGCTGCAGACGATGCCCTTGTCTCTTGATGCCCATGACCAAGCCGC  520

Db  615  TCTGGATTCTTCAGTATTTGTTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG  674
Qy  521  TGGTGGGCAAGCTGCCTCGCATGAGTACAGCAGCAGTGAGTCCTCGGGGCTTGCCCC  580

Db  675  AGGCTGAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTAAATCTCTTCCCT  734
Qy  581  TCATGGTCTCGTGGTGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGCGTGTCT  640

Db  735  GGATTCTGCTTGGGGCATGTTCATAGGATATGTACTTCCACTTAAATCATTCATCTCATCT  794
Qy  641  GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGACCAGCA  700

Db  795  GCTATTCTCAGATCTGCTCGGAACTCTTCAGAACTGCTGCCAAACAAAACCCACTCACTGAGA  854
Qy  701  GGAAAGGACGCCACTGGCCGAGGCTACGCCTGCTGCTCTGGGAATAGTTCTTATTGTTGGTTTGTTC  760

Db  855  AATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGGTGTGTTGTTTGTTC  914
Qy  761  TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG  801
              ||||
Db  915  TCTGTTTCACACCTTACCATGTTGCAATTATTCAACATATG  955
```

RESULT 6
BX417963
```
LOCUS       BX417963                1201 bp    mRNA     linear   EST 13-MAY-2003
DEFINITION  BX417963 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE014YH17
            5-PRIME, mRNA sequence.
ACCESSION   BX417963
VERSION     BX417963.1  GI:3063333
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6994.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DE014CD09QP1&cluster=6994.r  Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DE014CD09QP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DE014YH17"
                     /tissue type="PLACENTA"
                     /clone_lib="Homo sapiens PLACENTA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
```

```
Db  847  TCTGTTCACACCTTACCATGTTGCAATTATTCAACATATGATTAAGAAGCTTCGTTTCT  906
Qy  821  TGCCACTCCTGT------GCCGAGCGGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTG  873
              |||
Db  907  CTAATTTCCTGGAATGTAGCCAAAAGACATTCGTTYCAGATTTCTGCACTTTACAGTA  966
              |||||                           |||||
Qy  874  GCCCTCATGAACATGAACTGGCATTACCCCAATCATTACTTCTTTGCAT  925
              |||||| ||| |||       |||||  ::|:::||||||
Db  967  TGCCTGATGRACYTCAATTGSTGCATGGSTGCATGTGGGCCCYTTTATCWCYTYTTTGCAT  1018
```

RESULT 5
AL532537
```
LOCUS       AL532537                 979 bp    mRNA     linear   EST 23-MAY-2003
DEFINITION  AL532537 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM006YC06 5-PRIME, mRNA sequence.
ACCESSION   AL532537
VERSION     AL532537.2  GI:31070369
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12796030.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6994.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DM006BB03QP1&cluster=6994.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DM006BB03QP1.
FEATURES             Location/Qualifiers
     source          1..979
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM006YC06"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."

ORIGIN
Query Match     14.5%;  Score 160.2;  DB 9;  Length 979;
Best Local Similarity   51.8%;  Pred. No. 5.4e-23;
Matches  363;  Conservative   0;  Mismatches 338;  Indels   0;  Gaps   0;

Qy  101  CTCTGTCTCACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC  160
              ||||                                                   |||
Db  255  CCTGCATTACAGCCCTGCTCTTCATCATTGGGCTCGTGTGGGAAACTTACTAGCCTTCGTCG  314
Qy  161  TTACCTGTCAAAAGAGACAGGAGGAGATCAACTGCACAGGCATCTACCTGGTGACCGGCTG  220
              ||||   |||||||   |||    ||||||||
Db  315  TCATTGTTCAAAACAGGAGAAAAAATCAACTCTCTACCACCCTCTATTCAACAAATTTGGTGA  374
Qy  221  TGTCTGACCTGCGTCGTTCACCGTGCCTTACCGTGCGTCGGAGGGTCGTGTTATGTCGTCGGGCT  280
              |||||  |    |||||  ||||||  |||
Db  375  TTTCTGATATACTTTTTACCACCGGCTTTGCCTACACGAAATAGCCTATATGCAATGGGGCT  434
Qy  281  CCAGCTGCGTTCGGCAAGGGGCTCTCGCAGGCTGACGGCGTTTGTGCTCTACACCCACA  340
```

```
TITLE     A second set of porcine ESTs from a pooled-tissue normalized
          library
JOURNAL   Unpublished (2003)
COMMENT   Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68933-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@email.marc.usda.gov
          Single pass sequencing. Bases called with phred v0.020425.c and
          trimmed with the aid of the trim_alt option. Vector identified with
          cross_match v0.990329.
          Plate: SRG8011  row: N  column: 14
          Seq primer: TAGAAGGCACAGTCGAGG.
FEATURES             Location/Qualifiers
     source          1..636
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 3PIG"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with RNA pooled from multiple tissues
                     including brain, liver, muscle, placenta/endometrium,
                     ovary, testes, and bone marrow."
ORIGIN
```

```
Query Match        14.4%;    Score 159.4;   DB 14;   Length 636;
Best Local Similarity   54.8%;    Pred. No. 7.1e-23;
Matches  316;   Conservative    0;   Mismatches 261;   Indels    0;   Gaps    0;

Qy  101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
        ||||| ||||| ||||| ||||||||| ||||||||||||||||   ||||  |||||
Db  583 CTCTGCATTACAGCATCGTCTTCATATCGGGATCCTGGGAAACTTGTTGGCCCTGATTG 524

Qy  161 TTACCTGTCAAAGAGCAGGAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
        |||  |||||||||  |||||| |||||  ||||  ||||| |||| ||
Db  523 TCATTATTCAAAACAGGAAAAAATCAACTCAACTCTACCACTCTATATTCAACCAACTTGGTGA 464

Qy  221 TGTCTGACCTGCTGTTCACCGGGAAGGGTCGTGTTATGTGTCTGGGCT 280
        ||| ||||| | | |||| || | ||||
Db  463 TTTCGGATATACTGTTTACCACTGCTTCTGCCTACACGGATAGCCTACATGCGCTCGGGGCT 404

Qy  281 CCAGCTGGCCTTTCGGCAAGGGCTCTGCAGGCTGACGGCGTTTGTGCTTCACACCGACA 340
        |||| ||| ||||| | |||||  |||| || ||
Db  403 TTGACTGGAGAATTGGCGAGGCGGCCTTGTGCAGGCTGCAGGATAACTGCTCTCGTGTTTATATCAACA 344

Qy  341 CCTACGGGGGGGTCTACCTCATGCCTGTGTGACCGTGGACCATTACCCAGCTGTGGTCT 400
        ||||  || | |||||| |||||  |||| ||| || ||  |||||
Db  343 CCTACGGGGTCTACCTCATGGCCTGTGTGGCCGTGGAGCTGGACCATCCATTACCAGCTGTGGTCT 400

Qy  401 GTGCCCACTGGGGCCCCGCCCTCCGCACGGCCCCAGGCGTGGCTGCCA 460
        || || |||||| || || || | ||  || ||||  || ||  ||||
Db  401 GTGCCCACTGGGGCCCCGCCCTCCGCACGGCCCCAGGCGTGGCTGCCA 460

Qy  461 CTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGCCAAGCCGC 520
        ||||  || |||||||| || ||  | ||  |||| | |||
Db  461 AGGCTAAAAGGACTACGGTGCATGGAGTATCCAGAGTCCAAACTTTGAGGAAAACACCCTCCCT 520

Qy  521 TGGTGGGCAAGCTGGCCTGGCTGGCTGGCTGCTGATGGGAGTGGCAGCAGCATGGAGTCCTCCGGCTGCCCC 580
        || |||| ||||| |||| ||| |||| ||| ||  || |||| ||||
Db  521 TGGTGGGCAAGCTGGCCTGGCTGGCTGGCTGCTGATGGGAGTGGCAGCAGCATGGAGTCCTCCGGCTGCCCC 580

Qy  581 TCATGGTCCTGGTGGCCTTTCATAGGGTATGTACTTCCGCCTTAATCATCCTCTCATCT 640
        || | || | | || ||| ||
Db  581 TCATGGTCCTGGTGGCCTTTCATAGGGTATGTACTTCCGCCTTAATCATCCTCTCATCT 640

Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGC 677
        || || ||||||||||||| ||
Db   43 GCTACTCTCAGATCTGCTGCAAACTCTTCAAAACGGC 7
```

```
ORIGIN
          with a NotI-oligo(dT) primer. Five prime end enriched,
          double-strand cDNA was digested with Not I and cloned into
          the Not I and EcoRV sites of the pCMVSPORT 6 vector.
          Library was not normalized."

Query Match        14.5%;    Score 160.2;   DB 13;   Length 1201;
Best Local Similarity   51.8%;    Pred. No. 5.7e-23;
Matches  363;   Conservative    0;   Mismatches 338;   Indels    0;   Gaps    0;

Qy  101 CTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC 160
        ||||| |||||
Db  211 CTCTGCATTACAGCCTGCTCTTCATCATTGGGGAAACTTACTAGCCTTGGTCG 270

Qy  161 TTACCTGTCAAAGAGCAGGAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCTG 220
        |||
Db  271 TCATTGTTCAAAACAGGAAAAAATCAACTCAACCCTCTATTCAACAAATTTGGTGA 330

Qy  221 TGTCTGACCTGCTGTTCACCGGGAAGGGTCGTGTTATGTGTCTGGGCT 280
Db  271 TGTCTGACCTGCTGGCCTTCACCGGGAAGGGTCGTGTTATGTGTCTGGGCT 280

Qy  331 TTTCTGATATACTTTTTTACCACCGCCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 390
Db  331 TTTCTGATATACTTTTTTACCACCGCCTTTGCCTACACGAATAGCCTACTATGCAATGGGCT 390

Qy  281 CCAGCTGGCCAAGGGCTCTGCAGGGCTCGCAGGGCTGACGGCGTTTGTGCTTCTCACACCGACA 340
Db  281 CCAGCTGGCCAAGGGCTCTGCAGGGCTCGCAGGGCTGACGGCGTTTGTGCTTCTCACACCGACA 340

Qy  391 TTGACTGGAGAATCGGAGATGCCCTTGTTGTGAGGATAACTGCGCTAGTGTTTTACATCAACA 450
Db  391 TTGACTGGAGAATCGGAGATGCCCTTGTTGTGAGGATAACTGCGCTAGTGTTTTACATCAACA 450

Qy  341 CCTACGGGGGGGTCTACCTCATGCCTGTGTGACCGTGGACCGTGGACCATTACCCAGCTGTGGTCT 400
Db  451 CATATGCAGGTGTGAACTTATGACCTTGCCTGAGTATTGACCGGCTTCATTGCTGTGGTGC 510

Qy  401 GTGCCCACTGGGGCCCCGCCCTCCGCACGGCCCCAGGCGTGGCTGCCA 460
Db  401 GTGCCACTGGGGCCCCGCCCTCCGCACGGCCCCAGGCGTGGCTGGCCA 460

Qy  511 ACCCTCTACGCTACAAACAAGATAAAAAAGGATTGAACATGCAAAAGGCGTGTCATATTTG 570
Db  511 ACCCTCTACGCTACAAACAAGATAAAAAAGGATTGAACATGCAAAAGGGCGTGTCATATTTG 570

Qy  461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC 520
Db  461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCATGATGCGCTCTTGATGATGCCCATGACCAAGCCGC 520

Qy  571 TCATTATTCAGATTCTAGTATTTGCTCAGACACTCCCATATGTCTCCGGGCTCGCCC 630
Db  571 TCATTATTCAGATTCTAGTATTTGCTCAGACACTCCCATATGTCTCCGGGCTCGCCC 630

Qy  521 TGGTGGGCAAGCTGGCCTGGCTGGCTGGCTGCTGCATGCATGGAGTCCTCCGGCTGCCCC 580
Db  521 TGGTGGGCAAGCTGGCCTGGCTGGCTGGCTGCTGCATGCATGGAGTCCTCCGGCTGCCCC 580

Qy  631 AGGCTGAAAGGATTACATGCATGGAGTATCCAGAGTATCAAGAAACTTTGAAGAAACTCCTCCCT 690
Db  631 AGGCTGAAAGGATTACATGCATGGAGTATCCAGAGTATCAAGAAACTTTGAAGAAACTCCTCCCT 690

Qy  581 TCATGGTCCTGGTGGCCTTTCATAGGGTCTTCGTGTGGGCCAGTGGGGATCATCCTGTCCT 640
Db  581 TCATGGTCCTGGTGGCCTTTCATAGGGTCTTCGTGTGGGCCAGTGGGGATCATCCTGTCCT 640

Qy  691 GGATTCTGCTGGGCATGTTTCATAGGATAGTGTACTTCCACTTATAATCATTCTCATCT 750
Db  691 GGATTCTGCTGGGCATGTTTCATAGGATAGTGTACTTCCACTTATAATCATTCTCATCT 750

Qy  641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGTGTGCAGCACAGCA 700
Db  641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGTGTGCAGCACAGCA 700

Qy  751 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCTGCCAAAACAAACCCACTCACTGAGA 810
Db  751 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCTGCCAAAACAAACCCACTCACTGAGA 810

Qy  701 GGAAAGGAGCGCCACTGCGGAGGCGTCCCTGCTTACGGTGGGGCCGTGGTGG 760
Db  701 GGAAAGGAGCGCCACTGCGGAGGCGTCCCTGCTTACGGTGGGGCCGTGGTGG 760

Qy  811 AATCTGGTGTAAACAAAAGGCCTCCAACACAATTATTCTTATTATTGTTGTTGTGTTGTTC 870
Db  811 AATCTGGTGTAAACAAAAGGCCTCCAACACAATTATTCTTATTATTGTTGTTGTGTTGTTC 870

Qy  761 TCTGCTTCAGCCCTACCCCTACCACCTCAACATCAAGCAGTTCATG 801
Db  761 TCTGCTTCAGCCCTACCCCTACCACCTCAACATCAAGCAGTTCATG 801

Qy  871 TCTGTTTCACACCCTTACCATGTTGCAATTAATTGCAACATATG 911
Db  871 TCTGTTTCACACCCTTACCATGTTGCAATTAATTGCAACATATG 911
```

```
RESULT 7
CF359817/c
LOCUS       CF359817                636 bp    mRNA    linear   EST 25-AUG-2003
DEFINITION  821090 MARC 3PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION   CF359817
VERSION     CF359817.1  GI:34156919
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 636)
AUTHORS     Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
            Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
```

RESULT 8
AL551903
LOCUS   AL551903   1201 bp   mRNA   linear   EST 31-MAY-2003
DEFINITION AL551903 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI060YJ18 5-PRIME, mRNA sequence.
ACCESSION   AL551903
VERSION   AL551903.2 GI:31273719
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 15, 2001 this sequence version replaced gi:12890291.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6994.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI060DE09QP1&cluster=6994.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI060DE09QP1.
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI060YJ18"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match   14.4%; Score 159.4; DB 9; Length 1201;
Best Local Similarity 51.5%; Pred. No. 8.4e-23;
Matches 361; Conservative 2; Mismatches 338; Indels 0; Gaps 0;

Qy   101 CTTCGTTCTACACAGCCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCCTTTGCC 160
              |||||| || || | |||||| | |||| | || |||||||| ||||| |||
Db   226 CTCTCGCATTACAGCCTCGTCTTCATCATTGGGCTCGGTGGGAAACTTACTAGCCTTGGTCG 285

Qy   161 TTACCTGTCAAAAGAGCAGGAGAGATCAACTCACTGCACAGGCATCCACACTACCTGGCTG 220
              | || |||||||||||| |||| ||| ||| ||||||||||||| |||||||||||
Db   286 TCATTGTTCAAAACAGGAAAAAAATCAACTCTACCACCCCTCTATTCAACAAATTTGGTGA 345

Qy   221 TGTCTGACCTGCTGTTCACCGTGCCTTGGCCTTACCCGGAAGGGTCGTGTGTTATGTGCTG 280
              | ||||||||||| ||||||||||| |||||||||||||||||||| | ||||| ||
Db   346 TTTCTGATATACTTTTTTACCACCGTTTTACCACCGTGCCTTGCCTTGCCTTGCTTG 405

Qy   281 CCAGCTGGCCTTCGGCAAGGGGGCTCTGCAGGCTGACGGGCGTTTGTGCTCTACACCGACA 340
              |||| | ||||| ||||||||| |||||||| | || |||||||||||||||||||||
Db   406 TTGACTGAGAATCGGGAATGCCTTGTGTAGGATAAACTTGCGCYAGTGTTTTACATCAACA 465

Qy   341 CCTACGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
              ||||||| ||| |||| ||||||| ||||||||||||||||||||||||| ||||| |
Db   466 CATATGCAGGTGTGAACTTTATGACCTGCTCCCTGAGTATTGACCGGCTTCCATTGCTGTGGTGC 525

Qy   401 GTGCCCACTGGGGCCCGGCGCCCCGCCTCCGCACAGGCTGGCCGCGCGGGCTGGCTGTCGGCGTGGCCA 460
              |||||

Qy   526 ACCCTCTACGCTACAACAAGATAAAAGGATTGAACATGCAAAAGCGTGTGCATATTTG 585
              ||||

Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTCGTTGCTTGTGATGCCCATGACCAAGCCGC 520
              ||||||

Db   586 TCTCGGATTCTAGTATTCTCTCAGACACTCCCACTCCTCTCATCAACCCTATGTCAAAGCAGG 645

Qy   521 TGGTTGGGCAAGCTGGCGCCCGCATGGGCATGGAGGTACAGCACGCAGCATGGAGTGAGTACAGCAGCTGCCCC 580
              |||| ||||||||| ||||||| ||||||| |||||||||||||||||||||||||||

Db   646 AGGCTGAAAGGATTACATGCATGGAGTGAGTATCCAAACTTTGAAGAAACTAAAATCTTTCCCT 705

Qy   581 TCATGGTCCTGTGGTGCCTTGCCATTGCTTCTGTGTGGGCCAGTGGGGATCATCCTGTCCT 640
              ||| |||||||||| || |||||||||| |||||| |||

Db   706 GGATTCTGCTTGGGGCATGTTTCATAGGATGATGTACTTTCCACTTATAATCATTCTCATCT 765

Qy   641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCAGTGCAGCAGTCGGGAGGACCCAGTGACCACGCA 700
              ||||| |||||| |||||| ||| ||| || ||| |||||||||| ||||||||

Db   766 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCACTGAGA 825

Qy   701 GGAAAGGACGCCACTGGGCGAGGCGTGCCGTCTTACGCTGCTGATGGCTCGGTGGCCCGTGGTCGG 760
              |||||| || ||| || || ||| ||||||| |||||| |||||||||||||||

Db   826 AATCTGGTGTAAACAAAAAGGCTCTCAACACAATTATTCTTATTATTGTGTGTTTGTTC 885

Qy   761 TCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATG 801
              ||||||

Db   886 TCTGTTTCACACCTTACCCATGTTGCAWTTATTCAACATATG 926

RESULT 9
BM549799
LOCUS   BM549799   1088 bp   mRNA   linear   EST 20-FEB-2002
DEFINITION AGENCOURT_6543842 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5745818
            5', mRNA sequence.
ACCESSION   BM549799
VERSION   BM549799.1 GI:18785503
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1088)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12769 row: g column: 03
            High quality sequence stop: 619.
FEATURES   Location/Qualifiers
     source   1..1088
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5745818"
                     /tissue_type="leukocyte"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC 118"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
                     (destroyed); RNA source leukocytes from anonymous pool of
                     non-activated adult donors. Library is oligo-dT primed
                     and directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.7 kb, insert size range
                     1.2-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 027. Note:
                     this is a NIH_MGC Library."
ORIGIN

Query Match   14.1%; Score 155.6; DB 12; Length 1088;
Best Local Similarity 53.8%; Pred. No. 5.1e-22;
Matches 320; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

```
          Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
          Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
          Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
          Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
          Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
          Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
          Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
          Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
          Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
  JOURNAL Nature 420, 563-573 (2002)
  MEDLINE 22354683
   PUBMED 12466851
  COMMENT Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center(GSC) Yokohama Institute
          The Institute of Physical and Chemical Research (RIKEN)
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/
          Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
          Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
          Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
          Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
          Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
          Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
          Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
          Direct Submission
          Computational Analysis of Full-Length Mouse cDNAs Compared with
          Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. 10 (10), 1617-1630 (2000)
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res.
          10 (11), 1757-1771 (2000)
          Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
          Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
          Trust/MRC building Addenbrookes Hospital Cambridge) whose
          assistance we gratefully acknowledge.
          Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.
FEATURES             Location/Qualifiers
     source          1..715
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F83003E22"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched, activated spleen"
ORIGIN
Query Match     14.1%;  Score 155.2;  DB 13;  Length 715;
Best Local Similarity   52.9%;  Pred. No. 5.6e-22;
Matches   359;  Conservative    0;  Mismatches 313;  Indels    6;  Gaps      1;

Qy    293 TCGGCAAGGGGCTCTGACGGCGTTTGTCTACACCGACACCTACGGGGGGG  352
          ||  ||  ||| ||||| || |||||  || || |||||||| |||||||| |
Db     35 TAGGTGATGCCCTGCCCGGGTAACTGCCTCGTGTTTCATCAACACGTACGCAGGTG   94
Qy    353 TCTACCTCATGGCCGTGTGTGAGGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCACTGGG  412
```

```
Qy    101 CTCTGTTCTACACAGCCTCCTGGTCTTCAGTGCCTGGGAAACATCCTTGCCCTTTGCC  160
          ||| ||   |||
Db    190 CTCTGCATTACAGCCCTCGTCTTCATCATTGGGCTCGTGGGAAACTTACTAGCCTTGGTCG  249
Qy    161 TTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGGCATCTACCTGGTGCACCTGGCTG  220
          |||| |                 | |||   |  || || || | ||
Db    250 TCATTGTTCAAAACAGGAAAAAAAATCAACTCACCACCCCTCATTCAACAAATTGGTTA  309
Qy    221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGGAAGGGGTGGTGTTATGTGCTGGGCT  280
          | | |||| | |                      ||||||
Db    310 TTTCTGATATACTTTTTACCACCGCTTTGCCTACACGGAATAGCCTACTATGCAATGGGCT  369
Qy    281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGCAGCGCGTTTGTGCCTCACACCGACA  340
          |||                     ||||||
Db    370 TTGACTGGAGAATCGGAGATGCCCTTGTGTGTAGGATAACTGCGCTAGTGTTTTACATCAACA  429
Qy    341 CCTACGGGGGGGTCTACCTCATGGCCGTCGTGTGAGCGTGACCATTACCCAGTCTGGTCT  400
          |  | |       |||
Db    430 CATATGCAGGTGTGAACTTTATGACCGGCTTGCCTGAGTATTGACCGCTTCATTGCTGTGTGC  489
Qy    401 GTGCCCACTCGGGGCCCGCCTCCCGCACGGCCTGGCCGCCCAGGCTGGTCTGCGTGGCCA  460
          |||
Db    490 ACCCTCTCACGCTACACACAGATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG  549
Qy    461 TCTGGACCTTGGTGCTGCGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGC  520
          |||||  | |  |     || ||    |||
Db    550 TCTGGATTCTAGTATTTGCTCAGACACTCCATCCACTCCCATCTCAACCCATGTCAAAGCAGG  609
Qy    521 TGGTGGGCAAGCTGGCTGCATGGAGTACAGCAGCAGATGAGTCAGTCTCCGGCGCTGCCCC  580
          |||
Db    610 AGGCTGAAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAACTANATCCTTCCCT  669
Qy    581 TCATGGGTCCGTGGTGGCCTTTGCCATTGGCTTCGTGGGCCAGTGGGGATCATCCTGTTCCT  640
          |||
Db    670 GGATTCTGCTTGGGGCATGTTTCATAGGGATATGTACTTCCACTTATAATCATTCTCATCT  729
Qy    641 GCTATATGAAGATCACCTGGAAGCTGTGCAGCCAAGCTCGGGAGGAGACCCAGTGAC  695
          |||                    |||
Db    730 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGCCAAACAAAACCCACTCAC  784
```

RESULT 10
BY765663
LOCUS   BY765663   715 bp   mRNA   linear   EST 17-DEC-2002
DEFINITION BY765663 RIKEN full-length enriched, activated spleen Mus musculus
            cDNA clone F830033E22 3', mRNA sequence.
ACCESSION   BY765663
VERSION   BY765663.1 GI:27203835
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 715)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,

```
Db        |||| ||||| |||| |||| ||||
      95  TGAACTTCATGACTTGCTTGAGCATAGACCGCTTCTTCGCTGTGGTGCACCCTCTGCGCT  154

Qy   413  GCCGGCGCCTCCGCACGGCTGGCGCGCGGCCAGGCTGGTCTGCGTGCCATCTGGACCTTGG  472
          ||   |||
Db   155  ACAACAAGATTAAAGAATCGAATACGCAAAGGGTGTCTGCCTGCTCGTCTGGATTCTGG  214

Qy   473  TGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCATGACCAAGCCGCTGGTGGGCCAAGC  532
          ||| ||  |
Db   215  TCTTTGCTCAAACACTGCGCTGCTCCTCACCCCATGTCTAAGGAGGAGGGAGACAAGA  274

Qy   533  TGCCCTGCATGGAGTACAGCAGCAGCATGGAGTCGTCCGGGCTGCCCCTCATGGTCCTGG  592
          ||||||||||||||
Db   275  CCACTTGCATGGAGTATCCAAACTTTGAAGGACAGCAGGCGTCCTGCCGTGGATTCTGCTCG  334

Qy   593  TGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGCTCGTCTGCTATATGAAGA  652
          |||||||
Db   335  GAGCCTGTCTGCTGGGCTCACGTGCTGCCTATCACAGTGCATTCCTCGTGTTACTCTCAGA  394

Qy   653  TCACCTGGAAGCTGTGCAGCACAGCTCGGAGGAGCACCAGTGACCAGGAGGAAAGGACGCC  712
          || ||||
Db   395  TCTGCTGCAAACTCTTCAGGACTGCCAAGCACGAACCCACTCACCGAGAAATCTGGTGTGA  454

Qy   713  ACTGGGCGAGGCTGCCTGCGTACCGCTGCTGATGCTGGTGGCCCGTGGTCGTCTTCAGCC  772
          ||    |
Db   455  ACAAAAAGGCTCCTCCAACACAATTATCCTCATCATTGTGTCGTGTTCATCCTGTGCTTCACGC  514

Qy   773  CCTTACCACCTCAACATCAAGCAGTTCATGCGGCGAGAGGGATGCTCCACCTGCCATCC----  828
          ||| ||||||
Db   515  CCTACCACGTGGCCATCATTCAGCACATGATAAGATGCTCTCGTGCTCCCTGGAGCCCTGG  574

Qy   829  --TGTGCCGAGCGGAGGGGTTTCTTACTGTCCCTTCCAGGCCACCGTGCCCCTCATGAACA  886
            |||||||||||||
Db   575  AGTGTGGGGCGGCGAGACATTCCTTCCGAGATCTCTGCTCACGGTGTGCCTGATGATGAACT  634

Qy   887  TGAACTGTGGCATTACCCCAATCATTACTTCTTTGCATCCATCCCATTACAGGAAATGGC  946
          ||||| |||
Db   635  TCAACTGCTGCATGGACCCGTTCATATATTTCTTTGCATGCAAAGGGTATAAGAGAAAGG  694

Qy   947  TCCTGGGCATTTTAAAGC  964
          |||| |||
Db   695  TCATGAAGATGCTCAAAC  712
```

RESULT 11
BY745873
```
BY745873
LOCUS       BY745873               963 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY745873 RIKEN full-length enriched, 2 days neonate thymic
            cells (NOD) Mus musculus cDNA clone E430001G12 5', mRNA sequence.
ACCESSION   BY745873
VERSION     BY745873.1  GI:27172282
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 963)
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
```

Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,J., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,T.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.

```
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
MEDLINE     22354683
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/,
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,N.,
            Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
            Computational Analysis of Full-length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
            Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
            Trust/MRC building Addenbrookes Hospital Cambridge) whose
            assistance we gratefully acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..963
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="E430001G12"
                     /tissue_type="thymus"
                     /cell_type="thymic cells"
                     /clone_lib="RIKEN full-length enriched, 2 days neonate
                     thymus thymic cells (NOD)"
ORIGIN
```

```
Query Match          13.7%;  Score 151.2;  DB 13;  Length 963;
Best Local Similarity  55.5%;  Pred. No. 4.2e-21;
Matches  331;  Conservative  0;  Mismatches  263;  Indels  2;  Gaps  2;

Qy   101  CTCTGGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAACATCCTTGCCCTTTGCC  160
```

```
Db  170 CTCTGCATTACAGCCTGGTCTTCATCATTGGGCGTGGGAAACCTGCTGGGCCTTGGTTG  229
            |||| |||||    |||          |||              ||||
Qy  161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGCATCTACCTGGTGCACCTGGCTG    220
            |||| ||| | |  | |||  |    | ||   |  ||||  ||||| | |  ||
Db  230 TCATTGTTCAAAACAAGAAAAAAATCAACTCAACCACCTCTCTATTCAATGAACTTGGTCA  289
            |||  || || || ||     || |||||||  |||| | | |||| |||||| ||
Qy  221 TGTCTCGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT    280
            ||||| || |||||| ||||      |  | |||||||||||||| ||||||||||||
Db  290 TTTCTGACATCCTGTTTACCACAGCTTTACCCACTCCGGATAGCCTACATGCCTGGGCT    349
            |||||||      |||   || ||||  |||   || || |||   |||||||||||
Qy  281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGACGCGGTTTGTGCTTCTACACCGACA       340
            ||||||||||||  |||||||||||||| |  |||||||||  | ||||||||||
Db  350 TTGATTGGAGGATAGGTGATGGCCCCTGTGCCGGGTAACTGCTCTCGTGTGTTCTACATCAACA  409
            |||  ||| || || ||||  |||| ||| ||  |||||||  | ||||||||||||||
Qy  341 CCTACGGGGGGTCTACCTCATGTGACGCTGTGAGCGTGGACCATTACCCAGCTGTGGTCT   400
            ||||| |||||| |||| || || |  |||||| |||  ||| |  |  |||||||||
Db  410 CGTACGCAGGTGTGAACTTCATGACTTGCTGAGCATAGACCGCTTCGTGTCGTGTGC       469
            ||||   |||| ||  |||| ||||  || |||  |    |||||  ||| ||||||
Qy  401 GTGCCCACTGGGCCGCCCGCGCTGGCCTGGCCCAGGCTGGTCTGCGTCTGCGTCCGGCCA   460
            ||||||| | ||| || | || |||||| |  | |||| || |  ||  | ||  ||||
Db  470 ACCTCTGCGCTACAACAAGATT-AAAAGAATCGAATACGCAAAGGGTGTCTGCCTGTCCG    528
            |||| ||    ||  |||  ||  |||| || | |||  | |||||  ||| || ||||
Qy  461 TCTGGACCTTGGTGCTGCTGCGAGACGATGCCCTTGCTCTTGATGCCATGACCAAGCCGC   520
            ||||| ||| || |||||||   || || || || ||   |  |||| | |||||||
Db  529 TCTGGATTCCTGGTCTTTGCTCAAACCTGCTGCTCCTCACCCCATGTCTAAGGAGG      588
            |||||  |  || || | |||  ||| ||||||| || |||  ||||| |||||||
Qy  521 TGGTGGGCAAGCTGCCCTGCGATGGAGTACAGCCAGCCAGCAGCCGGATCATCCTGTCCT   580
            ||| ||  ||| |||  ||| |||||||| ||| || ||||||   |||| |||||
Db  589 AGGGAGACAAGACCACTTGCATGGAGTATCCAAACTTTGA-AGGGACAGCGTCCTGCCGT   647
            |||  |||||| | ||  |||||||||| ||| || ||  ||||||| |  ||||| |
Qy  581 TCATGTGTCCTGGTGGCTTTCTGTGGGCCAGTGGGGATCATCATCCTGTGTCCT        640
            |||||||||| |||||| | ||| ||||| |||||| |||| |||||| | |||
Db  648 GGATTCGTCTCGGAGCCTGTCTGTCGGGCTACGTTGCTGCCTATCACAGTCATTCTCCTGT  707
            ||||| |||| ||||| |||||||| |||  | || |  |||| ||||  ||||||||
Qy  641 GCTATATGAAGATCACCTGGAAGCGTGTGCAGCACAGCCAGTGACC               696
            |||||  |||| | ||| || |   || ||| ||||  |  | |||
Db  708 GTTACTTCCAGATCTGCTGCTGCCAAACTCTCCAGGACTGCCCAAGCAGACGAGACCCACTCACC  763
```

RESULT 12
BX427288

LOCUS	BX427288 1201 bp mRNA linear EST 15-MAY-2003
DEFINITION	BX427288 Homo sapiens FETAL LIVER Homo sapiens cDNA clone CS0DM009YF17 5-PRIME, mRNA sequence.
ACCESSION	BX427288
VERSION	BX427288.1 GI:30786637
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1201)
AUTHORS	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6994.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AM009CC09QP1&cluster=6994.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0AM009CC09QP1.
FEATURES	Location/Qualifiers
source	1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

          /clone="CS0DM009YF17"
          /tissue_type="FETAL LIVER"
          /dev_stage="fetal"
          /clone_lib="Homo sapiens FETAL LIVER"
          /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
          was primed with a NoI-oligo(dT) primer. Five prime end
          enriched, double-strand cDNA was digested with Not I and
          cloned into the Not I and EcoRV sites of the pCMVSPORT 6
          vector. Library was not normalized."

ORIGIN

Query Match   13.6%;   Score 150.6; DB 13; Length 1201;
Best Local Similarity   53.6%; Pred. No. 5.9e-21;
Matches 309; Conservative   2; Mismatches 266; Indels   0; Gaps   0;

```
Qy  101 CTCTGGTCTCACAGCCCTCCTTCGGTCTTCAGTGCCTGGGAAACATCCTGCCCTTTGCC      160
            |||| |||||||||| || ||||    || ||||    ||||  ||||| |||||| |||
Db  268 CTCTGCATTACAGCCCTCGTCTTCATCATTGGGCGTCCGTGGGGAAACTTACTAGCCTTGGTCG  327
Qy  161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGCATCACCTGGTGCACCTGGCTG        220
            |||| ||| || ||| ||||  |||||||||  |  |||    |  |||  | |
Db  328 TCATTGTTCAAAACAGGAAAARRATCAACTCAACCCTCTATTCAACAAATTGGTGA         387
Qy  221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTGGGCT       280
            ||||| |   | ||||  ||  | || | ||   | |  |  |  |||  |||||||||
Db  388 TTTCTGATATACTTTTTTACCACCCGCTTTGCCTACACCCCATGTCCTACCATGCAATGGGCT   447
Qy  281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCGGTTTGTGCTTCTCACACCGACA        340
            |||||||||||||  |||||||||||||| || |||||||||  | |  ||||||||
Db  448 TTGACTGGAGAATCGGAGATGCCCTTGTGTGAGGATAACTGCGCAGTGTTTTACATCAACA    507
Qy  341 CCTACGGGGGGTCTACCTCATGTGCCTGTGTGAGCGTGACCAGCTGTGGTCT             400
            ||||| |||||| |||| ||| || |||| |||| | ||||| | |  |||
Db  508 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATTGACCGCTTCATTGTGTGTGTGC     567
Qy  401 GTGCCCACTGGGCGCCCGCGCTGGCCTGGCCCAGGCTGGTCTGCGTCTGGTGGCCA         460
            ||||||| |  || || | || ||  || || |  | |  |  | ||  | |
Db  568 ACCCTCTACGCTACACAACAGATAAAGAGATTGAACATGCAAAAGGGTGTGCATATTTG      627
Qy  461 TCTGGACCTTGGTGCTGCTGCGAGACGATGCCCTTGCTCTTGATGCCATGACCAAGCCGC     520
            ||||| ||| || |||| || |||| ||| | ||| || |  |||||| |||||||
Db  628 TCTGGATTCAGATTATTGCTCAGACACTCCCACTCCTCATCAACCCTATGTCAAAGCAGG     687
Qy  521 TGGTGGGCAAGCTGCCGCTGATGGAGTACAGCAGCAGCATGGGATCAGTGAGTCCTGCCCC     580
            ||| ||||||| ||| |||||||||| || || || ||  | || |  | ||| || |||||
Db  688 AGGCTGAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAAACTAAATCTCTTCCT     747
Qy  581 TCATGTGTCCTGGTGGCCTTTGCCATTGGCTTTCTGTGGGCCCAGTGGGGATCATCCTGTCCT   640
            |||||||||| |||||| ||| |  | || ||||| ||||| ||| ||||| |||| |||
Db  748 GGATTCGTCTCGGGGCATGTTTCATAGGATAATGTACTTCCACTTATAATCATTCTCATCT    807
Qy  641 GCTATATGAAGATCACCTGGAAGCGTGTGCAGCACAGC                        677
            |||||  ||||| | || || || || ||| |  | ||
Db  808 GCTATTCTCAGATCTGCTGCAAACTCTTCAGAACTGC                         844
```

RESULT 13
AI659965/c

LOCUS	AI659965 594 bp mRNA linear EST 18-DEC-1999
DEFINITION	we64e01.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone IMAGE:2345880 3' similar to SW:EBI2_HUMAN P32249 EBV-INDUCED G PROTEIN-COUPLED RECEPTOR 2 ;, mRNA sequence.
ACCESSION	AI659965
VERSION	AI659965.1 GI:4763535
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE | 1 (bases 1 to 594)
AUTHORS | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1014   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 408.
FEATURES   Location/Qualifiers
     source   1..594
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:234580"
                     /dev_stage="fetal"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="Soares_thymus_NHFTh"
                     /note="Organ: thymus, pooled; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site 1: Not I;
                     Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAACGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo.   "

ORIGIN

Query Match   13.4%;   Score 148; DB 9;   Length 594;
Best Local Similarity 96.4%;   Pred. No. 1.7e-20;
Matches 162; Conservative   0; Mismatches   5; Indels   1; Gaps   1;

Qy   229 CTGCTGTTCACCTGGCCTTACCGGGAAGGGTGGTTGTTATGTGCTGGGCTTCCAGCTGG 288
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   590 CTGCTGTTCCACCTGGCCTTACCGGAAAGGGTGGTGTGTTATGTGCTGGGCTTCCAGCTGG 531

Qy   289 CCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTGTGCCTCCTACACCGACACCTACGGG 348
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   530 -CTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTGTGCCTCCTACACCGACACCTACGGG 472

Qy   349 GGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG 396
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   471 GGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCCGTG 424

RESULT 14
LOCUS   BY729793   662 bp   mRNA   linear   EST 17-DEC-2002
DEFINITION BY729793 RIKEN full-length enriched, adult male hippocampus Mus
            musculus cDNA clone C630037I05 5', mRNA sequence.
ACCESSION   BY729793
VERSION   BY729793.1 GI:27142920
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 662)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,

Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
MEDLINE   22354683
PUBMED   12466851
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
            Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..662
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="C630037I05"
                     /sex="male"
                     /tissue_type="hippocampus"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hippocampus"
                     /note="Site_1: XhoI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'

GAGAGAGAGAGGATCCAAGAGCGCTCTTTTTTTTTTTTTTTVN 3'), cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCCC 3']"

ORIGIN

Query Match   13.0%;   Score 143.4; DB 13;   Length 662;
Best Local Similarity 55.4%; Pred. No. 1.6e-19;
Matches 276; Conservative   0; Mismatches 222; Indels   0; Gaps   0;

Qy   101 CTCTGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
                ||||| |||||||||||||||| ||||| |||||||| |||||| |||||||||||
Db   165 CTCTGCATTACAGCCTGGTCTTCATCATTGGGCGTCGGTGGGAAACTTACTAGCCTTGGTTG 224

Qy   161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATTCACCTGGTCACCTGGCTG 220
                |||||| |||||||||||||| ||||| ||||||||||||||
Db   225 TCATTGTTCAAAACAGAAAAAAATCAACTCAACCACTCTCATTCAATGAACTTGGTCA 284

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAGAGGGTGGTGGTTATGTGTGGGCT 280
                |||||||||||||
Db   285 TTTCTGACATCGTGTTTACCACAGCTTTACCCACTCCGGATAGCCTACTATGCGGCTGGGCT 344

Qy   281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGTGACGGCGTTTGTGCTTCACACCGACA 340
                |||||
Db   345 TTGATTGGAGGATAGGTGATGCCCGTGCCGTAACTGCTCGTGGTGTTTCACATCAACA 404

Qy   341 CCTACGGGGGGGTCTACCTGCGGGTGTGATGCCGTGCTGGACCATTACCCAGCTGTGGTCT 400
                ||||
Db   405 CGTACGCAGGTGTGAACTTCATGACTTGCTTGAGCATAGACCGCTTCTTCGCTGTGGTGC 464

Qy   401 GTGCCCACTGGGGCCCGCGCCTCCGCACGGCTCGCCGCGCCCAGGCTGGTCTCGGGTGGCCA 460
                ||||||
Db   465 ACCCTCTGCCGCTACACAACAGATTAAAAGAATCGAATACGCAAAGGGTGTCTGCCTGTCCG 524

Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTTGATGCCCATGACCAAGCCGC 520
                ||||||
Db   525 TCTGGATTCGGTCTTTGCTCAAACACTGCCGTGCTCCTCACCCCTATGTCTAAGGAGAGG 584

Qy   521 TGGTGGGCAAGCTGGCCTGGCCTGCCATGAGTAGCACGCATGGAGTCAGTCCTCGGGCCCC 580
                |||||||||
Db   585 AGGGAGNACAGACCACTTGCATGGAGTATCCAAACTTTGAAGGGGACAGCGTCCTGCCGT 644

Qy   581 TCATGGTCCTGGTGGCCT 598
                ||||| ||||
Db   645 GGATTCTGCTCGCGGAGCCT 662

RESULT 15
BI768868
LOCUS   BI768868   861 bp   mRNA   linear   EST 25-SEP-2001
DEFINITION   603057774F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207011 5',
             mRNA sequence.
ACCESSION   BI768868
VERSION   BI768866.1 GI:15760446
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 861)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement:Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:

---

http://image.llnl.gov
Plate: LLAM11519 row: d column: 20
High quality sequence stop: 788.
FEATURES   Location/Qualifiers
  source   1..861
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:5207011"
               /clone_lib="NIH_MGC_122"
               /lab_host="DH10B"
               /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
               Site_1: NotI; Site 2: EcoRV (destroyed); RNA source
               anonymous pool of 24 week female lung, 16 week female
               spleen, and 20-22 week male spleens. Library is oligo-dT
               primed and directionally cloned (EcoRV site is destroyed
               upon cloning). Average insert size 1.4 kb, insert size
               range 1-3 kb. Library is normalized and enriched for
               full-length clones and was constructed by C. Gruber
               (Invitrogen)   Research Genetics tracking code 026. Note:
               this is a NIH_MGC Library."

ORIGIN

Query Match   12.7%;   Score 140.4; DB 12;   Length 861;
Best Local Similarity 54.3%; Pred. No. 7.5e-19;
Matches 305; Conservative   0; Mismatches 256; Indels   1; Gaps   1;

Qy   101 CTCTGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCC 160
                |||||||
Db   168 CTCTGCATTACAGCCTCGTCTTCATCATTGGGCGTCGGTGGGAAACTTACTAGCCTTGGTCG 227

Qy   161 TTACCTGTCAAAAGAGCAGGAGATCAACTGCACAGGCATCACCTGGTCACCTGGCTG 220
                |||||||
Db   228 TCATTGTTCAAAACAGAAAAAAATCAACTCAACCCTCTATTCAACAAATTGGTTA 287

Qy   221 TGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAGAGGGTGGTGTGTTATGTGTGGGCT 280
                |||||||
Db   288 TTTCTGATATACTTTTTACCACCGCTTTGCCACTCAGGTGCCTTGTGCTACTATGCAATGGGCT 347

Qy   281 CCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGTGACGGCGTTTGTGCTTCACACCGACA 340
                |||||||
Db   348 TTGACTGGAGAATCGGAGATCGCCTTGTAGGATAACTGCGCTAGTGTTTTACATCAAACA 407

Qy   341 CCTACGGGGGGGTCTACCTCATGGCCTGTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCT 400
                |||| ||||
Db   408 CATATGCAGGTGTGAACTTTATGACCTGCCTGAGTATGGACCGCTTCATTGCTGTGTGTGC 467

Qy   401 GTGCCCACTGGGGCCCGGACCTTCCGCACGGCTCGCCGCGCCCAGGCTGGTCTCGCGTGGCCA 460
                |||||||
Db   468 ACCCTCTACGCTACAACAAGATATAAAAAGGATTGAACATGCAAAAGGCGTGTGCATATTTG 527

Qy   461 TCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTTCTTGATGCCCATGACCAAGCCGC 520
                |||||||
Db   528 TCTGGATTCTAGTATTGCT-CAGACACACTTCCATCATCAACCCTATGTCAAAGCAGG 586

Qy   521 TGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCCTCGGGCTGCCCC 580
                |||||||
Db   587 AGGCTGAAAGGATTACATGCATGGAGTATCCAAACTTTGAAGAAAACTAAATCTCTTCCCT 646

Qy   581 TCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTTGTGGGGCCAGTGGGGATCATCCTGTCCT 640
                |||||||
Db   647 GGATTCTGCTTGGGGCATGTTTCATAGGATATGTGTACTTCCACTTATAATCATTCTCATCT 706

Qy   641 GCTATATGAAGATCACCTGGAA 662
                |||||||
Db   707 GCTATTCTCAGATCTGCTTGAA 728

Search completed: February 21, 2004, 21:26:24
Job time : 2800 secs

OM nucleic - nucleic search, using sw model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: February 21, 2004, 20:39:45 ; Search time 4338 Seconds
                              (without alignments)
                              11030.592 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1104
Sequence: 1 atgataaaactaggcccta.........gtgtcattgctttatgact 1104
Scoring table: DNARSONUC
                Gapop 60.0., Gapext 60.0
Searched: 3470272 seqs, 2167516995 residues
Word size: 10
Total number of hits satisfying chosen parameters: 2871005
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database :   GenEmbl:*
     1: gb_ba:*
     2: gb_htg:*
     3: gb_in:*
     4: gb_om:*
     5: gb_ov:*
     6: gb_pat:*
     7: gb_ph:*
     8: gb_pl:*
     9: gb_pr:*
    10: gb_ro:*
    11: gb_sts:*
    12: gb_sy:*
    13: gb_un:*
    14: gb_vi:*
    15: gb_ba:*
    16: em_fun:*
    17: em_hum:*
    18: em_in:*
    19: em_mu:*
    20: em_om:*
    21: em_or:*
    22: em_ov:*
    23: em_pat:*
    24: em_ph:*
    25: em_pl:*
    26: em_ro:*
    27: em_sts:*
    28: em_un:*
    29: em_vi:*
    30: em_htg_hum:*
    31: em_htg_inv:*
    32: em_htg_other:*
    33: em_htg_mus:*
    34: em_htg_pln:*
    35: em_htg_rod:*
    36: em_htg_mam:*
    37: em_htg_vrt:*
    38: em_sy:*
    39: em_htgo_hum:*
    40: em_htgo_mus:*
    41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1104	100.0	1104	6	AX702273	AX702273 Sequence
2		461	41.8	4319	6	AX647101	AX647101 Sequence
3		461	41.8	98861	9	AL136961	AL136961 Human DNA
4	c	368	33.3	151174	2	AC016189	AC016189 Homo sapi
5		176	15.9	536	6	AX230148	AX230148 Sequence
6	c	171	15.5	151174	2	AC016189	AC016189 Homo sapi
7	c	159	14.4	556	6	AX244735	AX244735 Sequence
8	c	34	3.1	147559	10	AC117702	AC117702 Mus muscu
9		34	3.1	242813	2	AC095386	AC095386 Rattus no
10	c	33	3.0	227	8	AY023453	AY023453 Oryza sat
11	c	33	3.0	1695	8	AK067772	AK067772 Oryza sat
12		33	3.0	2790	2	BT010460	BT010460 Arabidops
13		33	3.0	59159	2	AC100332	AC100332 Mus muscu
14		33	3.0	78244	2	AC140499	AC140499 Mus muscu
15		33	3.0	98286	2	AP004394	AP004394 Oryza sat
16		33	3.0	116540	2	AC138355	AC138355 Mus muscu
17		33	3.0	136949	10	AL845271	AL845271 Mouse DNA
18		33	3.0	144422	8	AC007258	AC007258 Arabidops
19		33	3.0	146963	2	AC138591	AC138591 Mus muscu
20	c	33	3.0	150495	2	AC046148	AC046148 Mus muscu
21		33	3.0	152515	10	AL928608	AL928608 Mouse DNA
22	c	33	3.0	159988	2	AP005823	AP005823 Oryza sat
23	c	33	3.0	170979	2	AC026270	AC026270 Homo sapi
24		33	3.0	171030	9	CNS01DSF	AL121774 Human chr
25		33	3.0	175377	2	AC120929	AC120929 Rattus no
26		33	3.0	181007	2	AC116729	AC116729 Mus muscu
27		33	3.0	184867	2	AC109147	AC109147 Mus muscu
28	c	33	3.0	188382	2	AC118696	AC118696 Mus muscu
29	c	33	3.0	191781	10	AC124765	AC124765 Mus muscu
30		33	3.0	193437	2	AC119524	AC119524 Rattus no
31		33	3.0	203409	2	AC123277	AC123277 Rattus no
32	c	33	3.0	205468	2	AC141491	AC141491 Rattus no
33		33	3.0	210156	2	AC141495	AC141495 Rattus no
34	c	33	3.0	210988	2	AC139449	AC139449 Rattus no
35		33	3.0	227492	10	AC116585	AC116585 Mus muscu
36		33	3.0	228380	2	AC133352	AC133352 Rattus no
37		33	3.0	248276	2	AC103007	AC103007 Rattus no
38		33	3.0	255296	2	AC108538	AC108538 Rattus no
39	c	33	3.0	255956	10	AC110237	AC110237 Mus muscu
40		33	3.0	259141	2	AC117864	AC117864 Rattus no
41		33	3.0	283450	2	AC127733	AC127733 Rattus no
42		33	3.0	286145	2	AC099222	AC099222 Rattus no
43		33	3.0	292748	2	AC106574	AC106574 Rattus no
44	c	33	3.0	332036	2	AC129629	AC129629 Rattus no
45		32	2.9	59698	2	AC116398	AC116398 Mus muscu

ALIGNMENTS

RESULT 1
AX702273
LOCUS   AX702273   1104 bp   DNA   linear   PAT 03-APR-2003
DEFINITION Sequence 33 from Patent WO02064793.
ACCESSION   AX702273
VERSION   AX702273.1 GI:29537547
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Casman,S.J., Edinger,S.R., Ellerman,K., Smithson,G., Kekuda,R. and
            Muralidhara,P.
TITLE   Novel gpcr-like proteins and nucleic acids encoding same

Pred. No. is the number of results predicted by chance to have a

```
Qy   721  GGCTGCCTGCTTACGCTGCTGATGCTGCTGATGCTGTGGCCGTGTGGGCCGTGGTCTGCTTGCTGCTTCAGCCCCTACCAC  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GGCTGCCTGCTTACGCTGCTGATGCTGCTGATGCTGTGGCCGTGTGGGCCGTGGTCTGCTTGCTGCTTCAGCCCCTACCAC  780

Qy   781  CTCAACATCAAGCAGTTCATTGGCGAGAGGATGTCCACCTGACTCCACCTGACTGCCGCCATCCTGTGCCGAGCGG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  CTCAACATCAAGCAGTTCATTGGCGAGAGGATGTCCACCTGACTCCACCTGACTGCCGCCATCCTGTGCCGAGCGG  840

Qy   841  AGGGCTTCTTACTGTCCCTTCAGGCACCGTGGCCCTCATGAACATGAACTGTGGCATT  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  AGGGCTTCTTACTGTCCCTTCAGGCACCGTGGCCCTCATGAACATGAACTGTGGCATT  900

Qy   901  ACCCCAATCATTACTTACTTCTTTGCATCCACCATTACAGGAAATGGCTCCTCGGCATTTTA  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  ACCCCAATCATTACTTACTTCTTTGCATCCACCATTACAGGAAATGGCTCCTCGGCATTTTA  960

Qy   961  AAGCTCAAAGGGTCTTCCTCTCCTCCTCCTCCTCCTCCTCCTCCCACCCCAGGAAAAGCTTCT  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AAGCTCAAAGGGTCTTCCTCTCCTCCTCCTCCTCCTCCTCCTCCCACCCCAGGAAAAGCTTCT  1020

Qy  1021  TCAGAAACACCAAGTATCACCCAGGCCAGGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  TCAGAAACACCAAGTATCACCCAGGCCAGGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC  1080

Qy  1081  TAACGTGTCATTGCTTTATGACT  1104
          |||||||||||||||||||||||
Db  1081  TAACGTGTCATTGCTTTATGACT  1104
```

RESULT 2
AX647101
LOCUS   AX647101   4319 bp   DNA   linear   PAT 04-MAR-2003
DEFINITION   Sequence 1293 from Patent EP1270724.
ACCESSION   AX647101
VERSION   AX647101.1   GI:28800143
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Asai,K., Akiyama,Y. and Aburatani,H.
  TITLE   Guanosine triphosphate-binding protein coupled receptors
  JOURNAL   Patent: EP 1270724-A 1293 02-JAN-2003;
               National Institute of Advanced Industrial Science and Technology
               (JP); Center for Advanced Science and Technology Incubation, Ltd.
               (JP)
FEATURES   Location/Qualifiers
     source   1..4319
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS   join(201..308,3128..3531,3660..4119)
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD69676.1"
                     /db_xref="GI:28800144"
                     /db_xref="REMTREMBL:CAD69676"
                     /translation="MFVDSVGQDIEBGTVGMIFLCSLTLEVSAEKDQQLGLAFRQGAL
                     QADGVCALHRLRGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQ
                     TMPLLMPWTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCTMKIT
                     WKLCSTAGRTQTAREDPVTSGKGRHRRDSPGPSDQQERTPLARLPAYAADAGGRGGL
                     LQPLPPHQAVHGERDAPPAILCRAEFGLTVPSGHRGPHEHELWHYPIIYFFASTHYR
                     KWLIGILKLKGSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV"
ORIGIN

Query Match   41.8%; Score 461; DB 6; Length 4319;
Best Local Similarity   99.8%; Pred. No. 3.5e-258;
Matches   581; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy   97  CGCTCTCGTTCTACACAGCCCCTCCGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  156
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

JOURNAL   Patent: WO 02064793-A 33 22-AUG-2002;
               Curagen Corporation (US)
FEATURES   Location/Qualifiers
     source   1..1104
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS   1..1083
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87717.1"
                     /db_xref="GI:29537548"
                     /db_xref="REMTREMBL:CAD87717"
                     /translation="MIKLGPYFTPTPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFS
                     ALGNILALCLTCQSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGL
                     CRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL
                     QTMPLLMPWTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKI
                     TWKLCSTAREDPVTSRGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPS
                     CAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSSS
                     TPGKASSETPSITQARGSMFLAEHVV"
ORIGIN

Query Match   100.0%; Score 1104; DB 6; Length 1104;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 1104; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy     1  ATGATAAAACTAGGCCCCTTACTTCACACACCCCACACCACCACCACCCAAAAATAAAAACTAAGATTAAAGAT  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGATAAAACTAGGCCCCTTACTTCACACACCCCACACCACCACCACCCAAAAATAAAAACTAAGATTAAAGAT  60

Qy    61  TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGCGCTCTCTGTTCTACACAGCCCTC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGCGCTCTCTGTTCTACACAGCCCTC  120

Qy   121  CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTGCCCTTTACCTGTCAAAAGAGCAGG  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTGCCCTTTACCTGTCAAAAGAGCAGG  180

Qy   181  AAGATCAACTGCACAGGACATTCACCTGGTGCTGCACCTGCTGTCTGACCTGCTGTTCACC  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  AAGATCAACTGCACAGGACATTCACCTGGTGCTGCACCTGCTGTCTGACCTGCTGTTCACC  240

Qy   241  GTGGCCTTACCGGGAAGGGTGGTGTGTTATGTCGGGGCTCCAGCTGCCCTTCGGCAAG  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GTGGCCTTACCGGGAAGGGTGGTGTGTTATGTCGGGGCTCCAGCTGCCCTTCGGCAAG  300

Qy   301  GGGGCTCTGCAGGCTGACGGCGTTTGTGCTTCTACACCGACACCTACCGGGGGGGGTCTACCTC  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GGGGCTCTGCAGGCTGACGGCGTTTGTGCTTCTACACCGACACCTACCGGGGGGGGTCTACCTC  360

Qy   361  ATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTTCTGTGCCCACTGGGGCCCGC  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  ATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTTCTGTGCCCACTGGGGCCCGC  420

Qy   421  CTCCGCACCGGCTGCGCGCCCAGGCTGGTCTGCGCTGGCCATCGGACCTGGTGCTGCTG  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  CTCCGCACCGGCTGCGCGCCCAGGCTGGTCTGCGCTGGCCATCGGACCTGGTGCTGCTG  480

Qy   481  CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCCTGC  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCCTGC  540

Qy   541  ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTGCCCCTCATGGTCCTGGTGGCCTTT  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  ATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTGCCCCTCATGGTCCTGGTGGCCTTT  600

Qy   601  GCCATTGGCTTCTGTGGCCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  GCCATTGGCTTCTGTGGCCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCACCTGG  660

Qy   661  AAGCTGTGCAGCACAGCTCGGAGGAGACCCAGTGACCAGCAGGAGAAAGGACGCCACTGGCGA  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  AAGCTGTGCAGCACAGCTCGGAGGAGACCCAGTGACCAGCAGGAGAAAGGACGCCACTGGCGA  720
```

```
Db   2940  CGCTCCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCGTGGGAAACATCCTTGCCCTT  2999
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    157  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG   216

Db   3000  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG  3059
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    217  GCTGTGTCTGCACCTGCTGTTCACCGTGGCCTTACCGGAAGGTGGTGTTGTTATGTGCTG   276

Db   3060  GCTGTGTCTGCACCTGCTGTTCACCGTGGCCTTACCGGAAGGTGGTGTTGTTATGTGCTG  3119
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    277  GGCTCCAGTTGGCCTTTCGGCAAGGGGCTCTGCAGCGGTGACGGCGGTTTGTGCTCTACACC   336

Db   3120  GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGCGGTGACGGCGGTTTGTGCTCTACACC  3179
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    337  GACACCTACGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG   396

Db   3180  GACACCTAC-GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTG  3238
                 |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    397  GTCTGTGCCCACTGGGGCCCGCGCCTCCCGCACGGCTGGCCGCCGGCCAGGCTGGTCTGCGTG   456

Db   3239  GTCTGTGCCCACTGGGGCCCGCGCCTCCCGCACGGCTGGCCGCCGGCCAGGCTGGTCTGCGTG  3298
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    457  GCCATCTGGACCTTGGTGTCGTGCGAGAGATGCCCTTGCTCTTGATGCCCATGACCAAG   516

Db   3299  GCCATCTGGACCTTGGTGTCGTGCGAGAGATGCCCTTGCTCTTGATGCCCATGACCAAG  3358
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    517  CGGCTGGTGGGCAAGCTGGCCTGCCATGGGCATGCAGCAGCAGCATGAGTCAGTCCTCGGGGCTG   576

Db   3359  CCGGCTGTGGGGCAAGCTGGCCTGCCATGGGCATGCAGCAGCAGCATGAGTCAGTCCTCGGGGCTG  3418
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    577  CCCCTCATGGTCCTGCGTGGCCTTTGCCATTGGCTTCTGTGTGGGCCAGTGGGGATCATCCTG   636

Db   3419  CCCCTCATGGTCCTGCGTGGCCTTTGCCATTGGCTTCTGTGTGGGCCAGTGGGGATCATCCTG  3478
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    637  TCCTGCTATATGAAGATCACCTGGAGCTGTGCAGCACAGCT   678

Db   3479  TCCTGCTATATGAAGATCACCTGGAGCTGTGCAGCACAGCT  3520
```

```
RESULT 3
AL136961/c
LOCUS       AL136961              98861 bp    DNA       linear   PRI 11-APR-2001
DEFINITION  Human DNA sequence from clone RP11-178C10 on chromosome 13,
            complete sequence.
ACCESSION   AL136961
VERSION     AL136961.19  GI:13620283
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98861)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 12, 2001 this sequence version replaced gi:13276997.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em, EMBL; Sw:,
```

```
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-178C10 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-178C10 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-178C10 is at 98861 in this
            sequence. The true right end of clone RP11-461N23 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..98861
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-178C10"
                     /clone_lib="RPCI-11.1"
     repeat_region   161..463
                     /note="L1MB8 repeat: matches 5836. .6160 of consensus"
     repeat_region   656..723
                     /note="L2 repeat: matches 2592. .2659 of consensus"
     repeat_region   2385. .2420
                     /note="MIR repeat: matches 111. .146 of consensus"
     repeat_region   4555. .4646
                     /note="L2 repeat: matches 1031. .1120 of consensus"
     repeat_region   4989. .5306
                     /note="MER7A repeat: matches 2. .345 of consensus"
     repeat_region   5859. .5987
                     /note="L2 repeat: matches 2574. .2708 of consensus"
     repeat_region   6215. .6548
                     /note="Charliel repeat: matches 2292. .2629 of consensus"
     repeat_region   6557. .6635
                     /note="L2 repeat: matches 2128. .2206 of consensus"
     repeat_region   6730. .7233
                     /note="L2 repeat: matches 1330. .1889 of consensus"
     repeat_region   8301. .8499
                     /note="MIR repeat: matches 17. .219 of consensus"
     repeat_region   8893. .8994
                     /note="L2 repeat: matches 1551. .1656 of consensus"
     repeat_region   9079. .9126
                     /note="24 copies 2 mer at 75% conserved"
     repeat_region   9554. .9754
                     /note="MER20 repeat: matches 33. .218 of consensus"
     repeat_region   9816. .9979
                     /note="FAM repeat: matches 2. .165 of consensus"
     repeat_region   10402. .10703
                     /note="L2 repeat: matches 5264. .5577 of consensus"
     repeat_region   12181. .12626
                     /note="L1M1 repeat: matches 3474. .3949 of consensus"
     repeat_region   12654. .13012
                     /note="Li repeat: matches 5571. .5927 of consensus"
     repeat_region   13491. .14076
                     /note="L1ME3 repeat: matches 2226. .2881 of consensus"
     repeat_region   16610. .17185
                     /note="Ul repeat: matches 6237. .7151 of consensus"
     repeat_region   17207. .17622
                     /note="L1MC4 repeat: matches 2. .413 of consensus"
     repeat_region   17528. .18380
                     /note="MER47B repeat: matches 7197. .7976 of consensus"
     repeat_region   20797. .21100
                     /note="L1MC4 repeat: matches 1. .297 of consensus"
     repeat_region   22746. .22814
                     /note="AluY repeat: matches 2574. .2643 of consensus"
     repeat_region   23758. .23981
                     /note="L2 repeat: matches 85. .310 of consensus"
                     /note="AluJb repeat:
```

repeat_region   25108...25212
                 /note="L2 repeat: matches 2570. .2688 of consensus"
repeat_region   25740...25795
                 /note="MIR repeat: matches 196. .251 of consensus"
repeat_region   25982...26012
                 /note="Tigger4 (Zombi) repeat: matches 2701. .2731 of consensus"
repeat_region   26005...26208
                 /note="Tigger4 (Zombi) repeat: matches 1772. .1970 of consensus"
repeat_region   26209...26513
                 /note="AluSx repeat: matches 1. .305 of consensus"
repeat_region   26514...28201
                 /note="Tigger4 (Zombi) repeat: matches 1. .1772 of consensus"
repeat_region   32616...32843
                 /note="6 copies 38 mer 86% conserved"
repeat_region   35533...35362
                 /note="15 copies 2 mer ac 100% conserved"
repeat_region   36220...36303
                 /note="2 copies 42 mer 97% conserved"
repeat_region   36967...37274
                 /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   39481...39767
                 /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region   41084...41111
                 /note="14 copies 2 mer aa 89% conserved"
repeat_region   41119...41174
                 /note="MLT1C repeat: matches 1. .58 of consensus"
repeat_region   41175...41482
                 /note="THE1C repeat: matches 1. .309 of consensus"
repeat_region   41483...41785
                 /note="AluSp repeat: matches 1. .309 of consensus"
repeat_region   41791...41842
                 /note="THE1C repeat: matches 58. .344 of consensus"
repeat_region   42051...42573
                 /note="L2 repeat: matches 1215. .1266 of consensus"
repeat_region   42857...43014
                 /note="MER74A repeat: matches 2. .555 of consensus"
repeat_region   43250...43279
                 /note="L2 repeat: matches 2553. .2733 of consensus"
repeat_region   44262...44708
                 /note="15 copies 2 mer ca 100% conserved"
repeat_region   44860...45175
                 /note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region   45204...45235
                 /note="MLT2FB repeat: matches 1. .328 of consensus"
repeat_region   45389...45559
                 /note="16 copies 2 mer tg 87% conserved"
repeat_region   45539...45788
                 /note="MER5B repeat: matches 4. .175 of consensus"
repeat_region   45963...46162
                 /note="MER5A repeat: matches 9. .184 of consensus"
repeat_region   46281...46837
                 /note="MIR repeat: matches 25. .250 of consensus"
repeat_region   47886...51275
                 /note="MER67D repeat: matches 1. .510 of consensus"
repeat_region   51364...51523
                 /note="L1PA6 repeat: matches 2744. .6141 of consensus"
repeat_region   51736...51791
                 /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region   51848...52167
                 /note="MIR repeat: matches 118. .178 of consensus"
repeat_region   55247...55308
                 /note="MSTB repeat: matches 86. .422 of consensus"
repeat_region   55309...55622
                 /note="MER69A repeat: matches 24. .87 of consensus"
repeat_region   55623...55715
                 /note="AluSq repeat: matches 1. .313 of consensus"
                 /note="MER69A repeat: matches 87. .176 of consensus"

repeat_region   57171...57271
                 /note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region   57462...57560
                 /note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region   57561...58049
                 /note="MER74A repeat: matches 1. .519 of consensus"
repeat_region   58063...58287
                 /note="AluSg/x repeat: matches 163. .300 of consensus"
repeat_region   58336...58475
                 /note="MLT1G repeat: matches 76. .205 of consensus"
repeat_region   58496...58809
                 /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   58818...59182
                 /note="L1ME3A repeat: matches 5789. .6157 of consensus"
repeat_region   59198...59336
                 /note="MER74A repeat: matches 1. .139 of consensus"
repeat_region   59337...59435
                 /note="MLT1G repeat: matches 169. .264 of consensus"
repeat_region   59626...59726
                 /note="MLT1F repeat: matches 437. .541 of consensus"
repeat_region   59804...60249
                 /note="L1ME3A repeat: matches 5301. .5782 of consensus"

Query Match   41.8%; Score 461; DB 9; Length 98861;
Best Local Similarity 99.8%; Pred. No. 4.1e-258;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Gaps 1;

Qy   97 CGCTCCTCTGTTTCTACACAGCCTCCTGGTCCTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56922 CGCTCCTCTGTTTCTACACAGCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 56863

Qy   157 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56862 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 56803

Qy   217 GCTGTGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTG 276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56802 GCTGTGTCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGTGGTGTTATGTGCTG 56743

Qy   277 GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTGTGCTCTACACC 336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56742 GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTGTGCTCTACACC 56683

Qy   337 GACACCTACGGGGGGGTCTACCTCATGGCCGGTGGAGCGTGGACCATTACCCAGCTGTG 396
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   56682 GACACCTAC-GGGGGGTCTACCTCATGGCCGTGGAGCGTGGACCATTACCCAGCTGTG 56624

Qy   397 GTCTGGTGCCCACTGGGGCCGCCTTCGCAGCAGCATGCCCTTGCTCTTGATGCCATGACCAAG 456
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56623 GTCTGTGCCCACTGGGGCCGCCTTCGCAGCAGCATGCCCGCCGCCAGGCTGGTCTGCGGTG 56564

Qy   457 GCCATCTGGACCTTGGTGCTGCTGCAGCAGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTG 516
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56563 GCCATCTGGACCTTGGTGCTGCTGCAGCAGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGGCTG 56504

Qy   517 CGGCTGGTGGGCAAGCTGGCCTGCAGCATGGAGTACAGCAGCATGGAGTCCTCGGGGCTG 576
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56503 CCGGCTGGTGGGCAAGCTGGCCTGCAGCATGGAGTACAGCAGCATGGAGTCCTCGGGGCTG 56444

Qy   577 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTG 636
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56443 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTG 56384

Qy   637 TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT 678
             ||||||||||||||||||||||||||||||||||||||||||
Db   56383 TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT 56342

RESULT 4
AC016189
LOCUS   AC016189   151174 bp   DNA   linear   HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016189

```
VERSION     AC016189.3  GI:9129134
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151174)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-23M9
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 151174)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrin,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6939915.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------ Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------ Project Information
            Center project name: L4399
            Center clone name: 23_M_9
            ------------------
          * NOTE: This record contains 152 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *        1     851: contig of  851 bp in length
          *      852     951: gap of  100 bp
          *      952    1804: contig of  853 bp in length
          *     1805    1904: gap of  100 bp
          *     1905    2719: contig of  815 bp in length
          *     2720    2819: gap of  100 bp
          *     2820    3668: contig of  849 bp in length
          *     3669    3768: gap of  100 bp
          *     3769    4657: contig of  889 bp in length
          *     4658    4757: gap of  100 bp
          *     4758    5600: contig of  843 bp in length
          *     5601    5700: gap of  100 bp
          *     5701    6532: contig of  832 bp in length
          *     6533    6632: gap of  100 bp
          *     6633    7481: contig of  849 bp in length
          *     7482    7581: gap of  100 bp
          *     7582    8447: contig of  866 bp in length
          *     8448    8547: gap of  100 bp
          *     8548    9383: contig of  836 bp in length
          *     9384    9483: gap of  100 bp
          *     9484   10342: contig of  859 bp in length
          *    10343   10442: gap of  100 bp
          *    10443   11323: contig of  881 bp in length
          *    11324   11423: gap of  100 bp
          *    11424   12254: contig of  831 bp in length
          *    12255   12354: gap of  100 bp
          *    12355   13288: contig of  934 bp in length
          *    13289   13388: gap of  100 bp
          *    13389   14235: contig of  847 bp in length
          *    14236   14335: gap of  100 bp
          *    14336   15172: contig of  837 bp in length
          *    15173   15272: gap of  100 bp
          *    15273   16130: contig of  858 bp in length
          *    16131   16230: gap of  100 bp
          *    16231   17073: contig of  843 bp in length
          *    17074   17173: gap of  100 bp
          *    17174   18003: contig of  830 bp in length
          *    18004   18103: gap of  100 bp
          *    18104   18939: contig of  836 bp in length
          *    18940   19039: gap of  100 bp
          *    19040   19885: contig of  846 bp in length
          *    19886   19985: gap of  100 bp
          *    19986   20834: contig of  849 bp in length
          *    20835   20934: gap of  100 bp
          *    20935   21761: contig of  827 bp in length
          *    21762   21861: gap of  100 bp
          *    21862   22743: contig of  882 bp in length
          *    22744   22843: gap of  100 bp
          *    22844   23708: contig of  865 bp in length
          *    23709   23808: gap of  100 bp
          *    23808   24676: contig of  868 bp in length
          *    24677   24776: gap of  100 bp
          *    24777   25602: contig of  826 bp in length
          *    25603   25702: gap of  100 bp
          *    25703   26560: contig of  858 bp in length
          *    26561   26660: gap of  100 bp
          *    26661   27504: contig of  844 bp in length
          *    27505   27604: gap of  100 bp
          *    27605   28449: contig of  845 bp in length
          *    28450   28549: gap of  100 bp
          *    28550   29398: contig of  849 bp in length
          *    29399   29498: gap of  100 bp
          *    29499   30357: contig of  859 bp in length
          *    30358   30457: gap of  100 bp
          *    30458   31323: contig of  866 bp in length
          *    31324   31423: gap of  100 bp
          *    31424   32300: contig of  877 bp in length
          *    32301   32400: gap of  100 bp
          *    32401   33254: contig of  854 bp in length
          *    33255   33354: gap of  100 bp
          *    33355   34218: contig of  864 bp in length
          *    34219   34318: gap of  100 bp
          *    34319   35169: contig of  851 bp in length
          *    35170   35269: gap of  100 bp
          *    35270   36125: contig of  856 bp in length
          *    36126   36225: gap of  100 bp
          *    36226   37061: contig of  836 bp in length
          *    37062   37161: gap of  100 bp
          *    37162   38003: contig of  842 bp in length
          *    38004   38103: gap of  100 bp
          *    38104   38945: contig of  842 bp in length
          *    38946   39045: gap of  100 bp
          *    39046   39882: contig of  837 bp in length
          *    39883   39982: gap of  100 bp
          *    39983   40833: contig of  851 bp in length
          *    40834   40933: gap of  100 bp
          *    40934   41779: contig of  846 bp in length
          *    41780   41879: gap of  100 bp
          *    41880   42722: contig of  843 bp in length
          *    42723   42822: gap of  100 bp
          *    42823   43678: contig of  856 bp in length
          *    43679   43778: gap of  100 bp
          *    43779   44634: contig of  856 bp in length
          *    44635   44734: gap of  100 bp
```

```
*  44735  45587: contig of 853 bp in length
*  45588  45687: gap of 100 bp
*  45688  46542: contig of 855 bp in length
*  46543  46642: gap of 100 bp
*  46643  47488: contig of 846 bp in length
*  47489  47588: gap of 100 bp
*  47589  48448: contig of 860 bp in length
*  48449  48548: gap of 100 bp
*  48549  49450: contig of 902 bp in length
*  49451  49550: gap of 100 bp
*  49551  50383: contig of 833 bp in length
*  50384  50483: gap of 100 bp
*  50484  51353: contig of 870 bp in length
*  51354  51453: gap of 100 bp
*  51454  52306: contig of 853 bp in length
*  52307  52406: gap of 100 bp
*  52407  53252: contig of 846 bp in length
*  53253  53352: gap of 100 bp
*  53353  54224: contig of 872 bp in length
*  54225  54324: gap of 100 bp
*  54325  55162: contig of 838 bp in length
*  55163  55262: gap of 100 bp
*  55263  56086: contig of 824 bp in length
*  56087  56186: gap of 100 bp
*  56187  57026: contig of 840 bp in length
*  57027  57126: gap of 100 bp
*  57127  58025: contig of 899 bp in length
*  58026  58125: gap of 100 bp
*  58126  59004: contig of 879 bp in length
*  59005  59104: gap of 100 bp
*  59105  59979: contig of 875 bp in length
*  59980  60079: gap of 100 bp
*  60080  60920: contig of 841 bp in length
*  60921  61020: gap of 100 bp
*  61021  61885: contig of 865 bp in length
*  61886  61985: gap of 100 bp
*  61986  62792: contig of 807 bp in length
*  62793  62892: gap of 100 bp
*  62893  63744: contig of 852 bp in length
*  63745  63844: gap of 100 bp
*  63845  64693: contig of 849 bp in length
*  64694  64793: gap of 100 bp
*  64794  65645: contig of 852 bp in length
*  65646  65745: gap of 100 bp
*  65746  66624: contig of 879 bp in length
*  66625  66724: gap of 100 bp
*  66725  67550: contig of 826 bp in length
*  67551  67650: gap of 100 bp
*  67651  68502: contig of 852 bp in length
*  68503  68602: gap of 100 bp
*  68603  69435: contig of 833 bp in length
```

```
Query Match          33.3%;  Score 368;  DB 2;  Length 151174;
Best Local Similarity 99.8%;  Pred. No. 1.7e-203;
Matches 488;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 1;

Qy     178 AGGAAGATCAACTGCACAGGCATCTACCTGGCTGTGCACCTGGCTGTCTGACCTGCTGTTC    237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138689 AGGAAGATCAACTGCACAGGCATCTACCTGGCTGTGCACCTGGCTGTCTGACCTGCTGTTC 138748

Qy     238 ACCGTGGCCTTACCGGGAAGGGGTGGTGTTATGTGCTGGGCTCCAGCGTGGCCTTTCGGC    297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138749 ACCGTGGCCTTACCGGGAAGGGGTGGTGTTATGTGCTGGGCTCCAGCGTGGCCTTTCGGC 138808

Qy     298 AAGGGGGCTCTGCGAGGCTGACGGCGGCGTTTGTGCTCTACACCGACAACTTACGGGGGGGTCTAC    357
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  138809 AAGGGGGCTCTGCGAGGCTGACGGCGGCGTTTGTGCTCTACACCGACAACACCTAC-GGGGGGTCTAC 138867

Qy     358 CTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCCCACTGGGGGCCCG    417
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138868 CTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGTGTCGTCTGTGCCCCACTGGGGGCCCG 138927

Qy     418 CGCCTCCGCACGGCTGGCCGCGCCAGGCTGGTCGTCTGGGCCATCTGGACCTTGGTGTGCTG    477
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138928 CGGCTCCGCACGGCTGCCGCCGGCGCACGGCTGGTCGTCTGGCCATCTGGACCTTGGTGTGCTG 138987

Qy     478 CTGCAGACGATGCCCTTGCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCC    537
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138988 CTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCC 139047

Qy     538 TGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGCGTCCTGTCCTCCTGTGGGC    597
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  139048 TGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGTCCTGTCCTCCTGTGGGGCCC 139107

Qy     598 TTTGGCCATTGGCTTCTTGTGGGCCAGTGGGATCATCCTGTCCTGTCTGCTGTATATGAAGATCACC    657
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  139108 TTTGGCCATTGGCTTCTTGTGGGCCAGTGGGATCATCCTGTCCTGTCTGCTGTATATGAAGATCACC 139167

Qy     658 TGGAAGCTG  666
           |||||||||
Db  139168 TGGAAGCTG 139176
```

```
RESULT 5
AX230148
LOCUS       AX230148                 536 bp    DNA        linear   PAT 11-SEP-2001
DEFINITION  Sequence 35 from Patent WO0162797.
ACCESSION   AX230148
VERSION     AX230148.1  GI:15592169
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Vogeli,G., Wood,L.S., Parodi,L.A. and Lind,P.
  TITLE     Novel g protein-coupled receptors
  JOURNAL   Patent: WO 016279-A 35 30-AUG-2001;
            PHARMACIA & UPJOHN COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..536
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match           15.9%;  Score 176;  DB 6;  Length 536;
Best Local Similarity 99.6%;  Pred. No. 6.3e-91;
Matches 226;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy     347 GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCC    406
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      69 GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGTGTCTGTGCCC 128

Qy     407 ACTGGGGGCCCGCGCCTTCGCACACGGCTGGCCGCGCACGGCTGGTCGTCTGGCCATCTGGA    466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 ACTGGGGGCCCGTGCCCGCGCCTTCGCACACGGCTGGCCGCGCACGGCTGGTCGTCTGGA 188

Qy     467 CCTTGGTCGTGCTGCGCAGCAGCATGCCCTGCTTCTTGATGCCCATGACCAAGCGCTGGG    526
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 CCTTGGTCGTGCTGCGCAGCAGCATGCCCTGCTTCTTGATGCCCATGACCAAGCGCTGGTGG 248

Qy     527 GCAAGCTGGCCTGCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG  573
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 GCAAGCTGGCCTGCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG 295
```

```
RESULT 6
AC016189/c
LOCUS       AC016189              151174 bp    DNA        linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-23M9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016189
VERSION     AC016189.3  GI:9129134
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 151174)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens, clone RP11-23M9
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 151174)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
           Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
           Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
           Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
           Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
           Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
           Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
           McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
           Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
           Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Jul 13, 2000 this sequence version replaced gi:6939915.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence submissions@genome.wi.mit.edu
           ------- Project Information
              Center project name: L4399
              Center clone name: 23_M_9
           -------
         * NOTE: This record contains 152 individual
         * sequencing reads that have not been assembled into
         * contigs. Runs of N are used to separate the reads
         * and the order in which they appear is completely
         * arbitrary. Low-pass sequence sampling is useful for
         * identifying clones that may be gene-rich and allows
         * overlap relationships among clones to be deduced.
         * However, it should not be assumed that this clone
         * will be sequenced to completion. In the event that
         * the record is updated, the accession number will
         * be preserved.
         *       1     851: contig of 851 bp in length
         *     852     951: gap of 100 bp
         *     952    1804: contig of 853 bp in length
         *    1805    1904: gap of 100 bp
         *    1905    2719: contig of 815 bp in length
         *    2720    2819: gap of 100 bp
         *    2820    3668: contig of 849 bp in length
         *    3669    3768: gap of 100 bp
         *    3769    4657: contig of 889 bp in length
         *    4658    4757: gap of 100 bp
         *    4758    5600: contig of 843 bp in length
         *    5601    5700: gap of 100 bp
         *    5701    6532: contig of 832 bp in length
         *    6533    6632: gap of 100 bp
         *    6633    7481: contig of 849 bp in length
         *    7482    7581: gap of 100 bp
         *    7582    8447: contig of 866 bp in length
         *    8448    8547: gap of 100 bp
         *    8548    9383: contig of 836 bp in length
         *    9384    9483: gap of 100 bp
         *    9484   10342: contig of 859 bp in length
         *   10343   10442: gap of 100 bp
         *   10443   11323: contig of 881 bp in length
         *   11324   11423: gap of 100 bp
         *   11424   12254: contig of 831 bp in length
         *   12255   12354: gap of 100 bp
         *   12355   13288: contig of 934 bp in length
         *   13289   13388: gap of 100 bp
         *   13389   14235: contig of 847 bp in length
         *   14236   14335: gap of 100 bp
         *   14336   15172: contig of 837 bp in length
         *   15173   15272: gap of 100 bp
         *   15273   16130: contig of 858 bp in length
         *   16131   16230: gap of 100 bp
         *   16231   17073: contig of 843 bp in length
         *   17074   17173: gap of 100 bp
         *   17174   18003: contig of 830 bp in length
         *   18004   18103: gap of 100 bp
         *   18104   18939: contig of 836 bp in length
         *   18940   19039: gap of 100 bp
         *   19040   19885: contig of 846 bp in length
         *   19886   19985: gap of 100 bp
         *   19986   20834: contig of 849 bp in length
         *   20835   20934: gap of 100 bp
         *   20935   21761: contig of 827 bp in length
         *   21762   21861: gap of 100 bp
         *   21862   22743: contig of 882 bp in length
         *   22744   22843: gap of 100 bp
         *   22844   23708: contig of 865 bp in length
         *   23709   23808: gap of 100 bp
         *   23809   24676: contig of 868 bp in length
         *   24677   24776: gap of 100 bp
         *   24777   25602: contig of 826 bp in length
         *   25603   25702: gap of 100 bp
         *   25703   26560: contig of 858 bp in length
         *   26561   26660: gap of 100 bp
         *   26661   27504: contig of 844 bp in length
         *   27505   27604: gap of 100 bp
         *   27605   28449: contig of 845 bp in length
         *   28450   28549: gap of 100 bp
         *   28550   29398: contig of 849 bp in length
         *   29399   29498: gap of 100 bp
         *   29499   30357: contig of 859 bp in length
         *   30358   30457: gap of 100 bp
         *   30458   31323: contig of 866 bp in length
         *   31324   31423: gap of 100 bp
         *   31424   32300: contig of 877 bp in length
         *   32301   32400: gap of 100 bp
         *   32401   33254: contig of 854 bp in length
         *   33255   33354: gap of 100 bp
         *   33355   34218: contig of 864 bp in length
         *   34219   34318: gap of 100 bp
         *   34319   35169: contig of 851 bp in length
         *   35170   35269: gap of 100 bp
         *   35270   36125: contig of 856 bp in length
         *   36126   36225: gap of 100 bp
         *   36226   37061: contig of 836 bp in length
         *   37062   37161: gap of 100 bp
         *   37162   38003: contig of 842 bp in length
         *   38004   38103: gap of 100 bp
         *   38104   38945: contig of 842 bp in length
         *   38946   39045: gap of 100 bp
         *   39046   39882: contig of 837 bp in length
         *   39883   39982: gap of 100 bp
         *   39983   40833: contig of 851 bp in length
         *   40834   40933: gap of 100 bp
         *   40934   41779: contig of 846 bp in length
         *   41780   41879: gap of 100 bp
         *   41880   42722: contig of 843 bp in length
         *   42723   42822: gap of 100 bp
         *   42823   43678: contig of 856 bp in length
         *   43679   43778: gap of 100 bp
         *   43779   44634: contig of 856 bp in length
         *   44635   44734: gap of 100 bp
         *   44735   45587: contig of 853 bp in length
         *   45588   45687: gap of 100 bp
         *   45688   46542: contig of 855 bp in length
         *   46543   46642: gap of 100 bp
```

```
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Vogeli,G. and Wood,L.S.
  TITLE     G protein-coupled receptors
  JOURNAL   Patent: WO 0166750-A 64 13-SEP-2001;
            PHARMACIA & UPJOHN COMPANY (US)
FEATURES            Location/Qualifiers
     source         1..556
                    /organism="Homo sapiens"
                    /mol_type="unassigned DNA"
                    /db_xref="taxon:9606"
ORIGIN

Query Match             14.4%;  Score 159; DB 6;  Length 556;
  Best Local Similarity 99.5%;  Pred. No. 6e-81;
  Matches  209;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy   697 AGCAGGAAAGACGACCACTGGCGAGGTGCCTGCTTACGCTGCTGATGCTGGTGGTGGCCGTG 756
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 AGCAGGAAAGACGACCACTGGCGAGGTGCCTGCTTACGCTGCTGATGCTGGTGGTGGCCGTG 497

Qy   757 GTGGTCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGCGAGAGGATGCTC 816
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 GTGGTCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGGCGAGAGGTATGCTC 437

Qy   817 CACCTGCCATCCTGTGCGAGCGGAGGGCTTCTTACTGTCCCTTCAGGCCACCGTGGCC 876
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 CACCTGCCATCCTGTGCGAGCGGAGGGCTTCTTACTGTCCCTTCAGGCCACCGTGGCC 377

Qy   877 CTCATGAACATGAACTGTGGCATTACCCCA 906
         |||||||||||||||||||||||||||||
Db   376 CTCATGAACATGAACTGTGGCATTACCCCA 347


RESULT 8
AC117702/c
LOCUS       AC117702             147559 bp    DNA     linear   ROD 04-DEC-2003
DEFINITION  Mus musculus chromosome 10, clone RP23-464I2, complete sequence.
ACCESSION   AC117702
VERSION     AC117702.10  GI:38678612
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 147559)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 10, clone RP23-464I2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147559)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,C., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Leboczky,J., Levine,J., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,R., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

```
*   46643   47488: contig of 846 bp in length
*   47489   47588: gap of 100 bp
*   47589   48448: contig of 860 bp in length
*   48449   48548: gap of 100 bp
*   48549   49450: contig of 902 bp in length
*   49451   49550: gap of 100 bp
*   49551   50383: contig of 833 bp in length
*   50384   50483: gap of 100 bp
*   50484   51353: contig of 870 bp in length
*   51354   51453: gap of 100 bp
*   51454   52306: contig of 853 bp in length
*   52307   52406: gap of 100 bp
*   52407   53252: contig of 846 bp in length
*   53253   53352: gap of 100 bp
*   53353   54224: contig of 872 bp in length
*   54225   54324: gap of 100 bp
*   54325   55162: contig of 838 bp in length
*   55163   55262: gap of 100 bp
*   55263   56086: contig of 824 bp in length
*   56087   56186: gap of 100 bp
*   56187   57026: contig of 840 bp in length
*   57027   57126: gap of 100 bp
*   57127   58025: contig of 899 bp in length
*   58026   58125: gap of 100 bp
*   58126   59004: contig of 879 bp in length
*   59005   59104: gap of 100 bp
*   59105   59979: contig of 875 bp in length
*   59980   60079: gap of 100 bp
*   60080   60920: contig of 841 bp in length
*   60921   61020: gap of 100 bp
*   61021   61885: contig of 865 bp in length
*   61886   61985: gap of 100 bp
*   61986   62792: contig of 807 bp in length
*   62793   62892: gap of 100 bp
*   62893   63744: contig of 852 bp in length
*   63745   63844: gap of 100 bp
*   63845   64693: contig of 849 bp in length
*   64694   64793: gap of 100 bp
*   64794   65645: contig of 852 bp in length
*   65646   65745: gap of 100 bp
*   65746   66624: contig of 879 bp in length
*   66625   66724: gap of 100 bp
*   66725   67550: contig of 826 bp in length
*   67551   67650: gap of 100 bp
*   67651   68502: contig of 852 bp in length
*   68503   68602: gap of 100 bp
*   68603   69435: contig of 833 bp in length

Query Match             15.5%;  Score 171; DB 2;  Length 151174;
  Best Local Similarity 100.0%;  Pred. No. 7.3e-88;
  Matches  171;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    907 ATCATTACTTCTTTGCATCCACCATTACAGGAAAATGGCTTCCTGGGCATTTTAAAGCTC 966
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 107201 ATCATTACTTCTTTGCATCCACCATTACAGGAAAATGGCTTCCTGGGCATTTTAAAGCTC 107142

Qy    967 AAAGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCCACCCCAGGAAAAAGCTTCTTCAGAA 1026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 107141 AAAGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCCACCCCAGGAAAAAGCTTCTTCAGAA 107082

Qy   1027 ACACCAAGTATCACCCAGGCCAGAGGTTCTATGTTCTTAGGCCGAGCACGTG 1077
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 107081 ACACCAAGTATCACCCAGGCCAGAGGTTCTATGTTCTTAGGCCGAGCACGTG 107031


RESULT 7
AX244735/c
LOCUS       AX244735               556 bp    DNA     linear   PAT 28-SEP-2001
DEFINITION  Sequence 64 from Patent WO0166750.
ACCESSION   AX244735
VERSION     AX244735.1  GI:15859614
KEYWORDS
SOURCE      Homo sapiens (human)
```

```
TITLE
JOURNAL    Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
           Direct Submission
           Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  3 (bases 1 to 147559)
AUTHORS    Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
           Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
           Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL    Direct Submission
           Submitted (07-NOV-2003) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE  4 (bases 1 to 147559)
AUTHORS    Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
           Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
           Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
           Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
           Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
           Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
           Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
           Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
           Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
           Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
           Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
           Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
           Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
           Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
           Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
           Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
           Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL    Direct Submission
           Submitted (04-DEC-2003) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Dec 4, 2003 this sequence version replaced gi:38198681.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -----------  Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           -----------  Project Information
           Center project name: L24127
           Center clone name: 464_I_2
           -----------
           Only the last 147.6 kilobases of this clone are being submitted.
           The remainder overlaps accession number AC13207 [WICGR project
           L23741].
FEATURES             Location/Qualifiers
     source          1..147559
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /map="10"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     repeat_region   1..31
                     /rpt_family="(GAGAA)n"
     repeat_region   complement(1556..1580)
                     /rpt_family="B3A"
     repeat_region   1581..1643
                     /rpt_family="(TG)n"
     repeat_region   complement(1644..1813)
                     /rpt_family="B3A"
     repeat_region   2382..2577
                     /rpt_family="B2_Mm2"
     repeat_region   complement(2801..3145)
                     /rpt_family="MTE"
     repeat_region   3528..3552
                     /rpt_family="(CA)n"
     repeat_region   3993..4358
                     /rpt_family="ORR1B2"
     repeat_region   4381..4524
                     /rpt_family="B1_MM"
     repeat_region   4526..4564
                     /rpt_family="(CAAAAA)n"
     repeat_region   4624..4798
                     /rpt_family="B2_Mm1"
     repeat_region   4866..4970
                     /rpt_family="B2_Mm2"
     repeat_region   7159..7327
                     /rpt_family="MTE"
     repeat_region   7477..7874
                     /rpt_family="RMER5"
     repeat_region   8257..8534
                     /rpt_family="B4A"
     repeat_region   complement(8583..8655)
                     /rpt_family="Lx5"
     repeat_region   10669..10834
                     /rpt_family="L1MB1"
     repeat_region   complement(11322..11439)
                     /rpt_family="PB1D10"
     repeat_region   complement(11584..11694)
                     /rpt_family="B3A"
     repeat_region   11906..12188
                     /rpt_family="B4A"
     repeat_region   13233..13323
                     /rpt_family="(CAAA)n"
     repeat_region   13660..13700
                     /rpt_family="(TCTG)n"
     repeat_region   complement(13791..14056)
                     /rpt_family="MLT2B2"
     repeat_region   15515..15535
                     /rpt_family="AT_rich"
     repeat_region   16640..17031
                     /rpt_family="Lx5"
     repeat_region   17222..17342
                     /rpt_family="ID_B1"
     repeat_region   17346..17594
                     /rpt_family="(GA)n"
     repeat_region   17763..18117
                     /rpt_family="MLT1A"
     repeat_region   18328..18360
                     /rpt_family="(CA)n"
     repeat_region   18360..18405
                     /rpt_family="(GA)n"
     unsure          18541..18570
                     /note="single clone coverage"
     repeat_region   18714..19054
                     /rpt_family="HAL1"
     repeat_region   complement(19277..19791)
                     /rpt_family="MLT2E"
     repeat_region   20505..20714
                     /rpt_family="B3"
```

```
repeat_region      complement(21076..21191)
                   /rpt_family="B1F"
repeat_region      complement(21193..21334)
                   /rpt_family="RIM"
repeat_region      21349..21383
                   /rpt_family="AT_rich"
repeat_region      complement(21472..21869)
                   /rpt_family="L1_MM"
repeat_region      21869..22286
                   /rpt_family="L1_MM"
repeat_region      22568..22668
                   /rpt_family="MT2B"
repeat_region      22595..22732
                   /rpt_family="B1_MM"
repeat_region      22738..22785
                   /rpt_family="GA-rich"
repeat_region      23525..23572
                   /rpt_family="(CA)n"
repeat_region      24142..24221
```

```
Query Match        3.1%;  Score 34;  DB 10;  Length 147559;
Best Local Similarity 100.0%;  Pred. No. 1.9e-07;
Matches   34;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     969 AGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
           |||||||||||||||||||||||||||||||||||||
Db   94177 AGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 94144
```

```
RESULT 9
ACCESSION   AC095386
LOCUS       AC095386           242813 bp    DNA    linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-601S, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC095386
VERSION     AC095386.8  GI:30467829
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242813)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,R., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,M.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarrpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rockey,T., Rojas,M., Rose,M., Rose,M., Ruiz,S.J.,
            Sanders,W., Savery,G., Scott,G., Shatsman,H., Scherer,S.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,R., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Vaisa.R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242813)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242813)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24941437.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------------------ Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            ------------------ Project Information
               Center project name: GCAQ
               Center clone name: CH230-601S
            ------------------ Summary Statistics
               Assembly program: Atlas;
               Consensus quality: 236293 bases at least Q40
               Consensus quality: 237786 bases at least Q30
               Consensus quality: 239195 bases at least Q20
               Estimated insert size: 248197; sum-of-contigs estimation
               Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *       consists of 4 contigs. The true order of the pieces
            *       is not known and their order in this sequence record is
            *       arbitrary. Gaps between the contigs are represented as
            *       runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *    1  237255: contig of 237255 bp in length
```

```
          *      237256   237355: gap of unknown length                                        /rpt_type=tandem
          *      237356   238503: contig of 1148 bp in length                                   /rpt_unit="tcc"
          *      238504   238603: gap of unknown length
          *      238604   239798: contig of 1195 bp in length             ORIGIN
          *      239799   239898: gap of unknown length
          *      239899   242813: contig of 2915 bp in length.            Query Match         3.0%;   Score 33;   DB 8;   Length 227;
     FEATURES         Location/Qualifiers                                 Best Local Similarity  100.0%;  Pred. No. 5.3e-07;
          source          1..242813                                       Matches   33;  Conservative   0;  Mismatches   0;   Indels   0;   Gaps   0;
                          /organism="Rattus norvegicus"
                          /mol_type="genomic DNA"                         Qy    973  TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC  1005
                          /db_xref="taxon:10116"                                    ||||||||||||||||||||||||||||||||
                          /clone="CH230-6O15"                             Db     98  TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC   130
          misc_feature    1..2171
                          /note="wgs end extension
                          clone_end:Sp6"                                   RESULT 11
          misc_feature    3006..3795                                      AK067772/c
                          /note="clone_boundary                           LOCUS       AK067772            1695 bp    mRNA        linear   PLN 24-JUL-2003
                          clone_end:Sp6                                    DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J013120L23, full
                          site:EcoRI                                                   insert sequence.
                          end_sequence:BH271023"                          ACCESSION   AK067772
          misc_feature    17728..189718                                   VERSION     AK067772.1  GI:32977790
                          /note="clone_boundary                           KEYWORDS    FLI cDNA; CAP trapper.
                          clone_end:T7                                     SOURCE      Oryza sativa (japonica cultivar-group)
                          site:EcoRI                                       ORGANISM    Oryza sativa (japonica cultivar-group)
                          end_sequence:BH271022"                                      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          misc_feature    23429..237255                                              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                          /note="wgs end extension                                    Ehrhartoideae; Oryzeae; Oryza.
                          clone_end:T7"                                    REFERENCE   1
                                                                          AUTHORS     The Rice Full-Length cDNA Consortium, National Institute of
     ORIGIN                                                                           Agrobiological Sciences Rice Full-Length cDNA Project Team.,
                                                                                      Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
     Query Match         3.1%;   Score 34;   DB 2;   Length 242813;                  Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
     Best Local Similarity  100.0%;  Pred. No. 2e-07;                                Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
     Matches   34;  Conservative   0;  Mismatches   0;   Indels   0;   Gaps   0;     Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
                                                                                      Science Genome Sequencing & Analysis Group., Otomo,Y., Murakami,K.,
     Qy    972  GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC  1005                              Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
                ||||||||||||||||||||||||||||||||||                                    Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
     Db  49752  GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC  49785                             Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
                                                                                      Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
     RESULT 10                                                                        Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN.,
     AY023453                                                                         Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
     LOCUS       AY023453             227 bp    DNA        linear   PLN 07-FEB-2001   Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
     DEFINITION  Oryza sativa microsatellite MRG5778 containing (TCC)X9, genomic      Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
                 sequence.                                                            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
     ACCESSION   AY023453                                                             Yoshino,M. and Hayashizaki,Y.
     VERSION     AY023453.1  GI:12706669                                 TITLE       Collection, mapping, and annotation of over 28,000 cDNA clones from
     KEYWORDS    .                                                                    japonica rice
     SOURCE      Oryza sativa                                            JOURNAL     Science 301 (5631), 376-379 (2003)
     ORGANISM    Oryza sativa                                           MEDLINE     22752273
                 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  PUBMED      12869764
                 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;          REFERENCE   2  (bases 1 to 1695)
                 Ehrhartoideae; Oryzeae; Oryza.                          AUTHORS     Adachi,J., Aizawa,T., Arakawa,T., Carninci,P., Doi,K.,
     REFERENCE   1  (bases 1 to 227)                                                 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
     AUTHORS     Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.                 Hayashida,Y., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
     TITLE       Simple sequence repeats from Monsanto rice genomic sequences        Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
     JOURNAL     Unpublished                                                         Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
     REFERENCE   2  (bases 1 to 227)                                                 Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
     AUTHORS     Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.                 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
     TITLE       Direct Submission                                                   Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
     JOURNAL     Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh     Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
                 Blvd., Creve Coeur, MO 63167, USA                                   Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
     COMMENT     Derived from rice genomic sequences generated from the Monsanto     Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
                 Rice Genome Sequencing project. Please see                          Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
                 http://www.rice-research.org for more information. The sequence     Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
                 data were produced primarily in the laboratories of Dr. Leroy Hood  Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saito,R., Sakai,C., Sakai,K.,
                 at the University of Washington in Seattle.                         Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
     FEATURES         Location/Qualifiers                                            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
          source          1..227                                                     Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
                          /organism="Oryza sativa"                                   Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
                          /mol_type="genomic DNA"                                    Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
                          /db_xref="taxon:4530"                                      Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
          repeat_region   1..227                                                     Yoshimura,A.
                          /note="microsatellite MRG5778                   TITLE       Direct Submission
                                                                          JOURNAL     Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
```

COMMENT   Agrobiological Sciences, Department of Molecular Genetics, Head of
           Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
           305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp;
           Tel:81-29-838-7007, Fax:81-29-838-7007)
           This clone is one of the 28K full-length cDNA clones from japonica
           rice.
           URL :http://cdna01.dna.affrc.go.jp/cDNA/
           NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
           Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
           Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., and
           Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
           Yamamoto,M.
           FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
           Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
           Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
           Mizuno,K., Narikawa,R., Nikura,J., Oka,M., Ryu,R., Sugano,S.,
           Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
           Yoshimura,A., Matsubara,K. and Murakami,K.
           Genome Exploration Research Group in Riken Genomic Sciences Center
           and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
           Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
           Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
           Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
           Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
           Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
           Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
           Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
           Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
           Yasunishi,A. and Hayashizaki,Y.

FEATURES
     source   Location/Qualifiers
               1..1695
               /organism="Oryza sativa (japonica cultivar-group)"
               /mol_type="mRNA"
               /cultivar="Nipponbare"
               /db_xref="taxon:39947"
               /clone="J013120L23"

ORIGIN

Query Match   3.0%;   Score 33;   DB 8;   Length 1695;
Best Local Similarity 100.0%; Pred. No. 5.9e-07;
Matches   33;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
            ||||||||||||||||||||||||||||||||||||
Db   230 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 198

RESULT 12
BT010460/c
LOCUS   BT010460   2790 bp   mRNA   linear   PLN 26-AUG-2003
DEFINITION   Arabidopsis thaliana formin homology 2 (FH2)
             domain-containingprotein   (At1g59910) mRNA, complete cds.
ACCESSION   BT010460
VERSION   BT010460.1   GI:34222087
KEYWORDS   FLI_CDNA.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 2790)
AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
             Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
             Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
             Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
             Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
             Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
             and Ecker,J.R.
TITLE   Arabidopsis ORF clones
JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 2790)
AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P.,
             Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
             Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
             Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
             Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
             Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A.
             and Ecker,J.R.
TITLE   Direct Submission
JOURNAL   Submitted (26-AUG-2003) Salk Institute Genomic Analysis Laboratory
             (SIGnAL), Plant Biology Laboratory, The Salk Institute for
             Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
             USA
COMMENT   RIKEN Genomic Sciences Center (GSC) members carried out the
             collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
             Arabidopsis Full-Length cDNA') :   Seki,M., Narusaka,M., Ishida,J.,
             Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
             Hayashizaki,Y. and Shinozaki,K.

             The Salk, Stanford, PGEC (SSP) Consortium members constructed and
             sequenced the pUNI (ORF) clones using the RAFL cDNAs: Shinn,P.,
             Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Chan, M.M., Chang,C.M.,
             Dale,J.M., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J.,
             Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
             Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
             Yu,G., Davis,R.W., Theologis,A., and Ecker,J.R.

             Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
             this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
             contributed equally to this work as PIs.
FEATURES   Location/Qualifiers
     source   1..2790
               /organism="Arabidopsis thaliana"
               /mol_type="mRNA"
               /db_xref="taxon:3702"
               /chromosome="1"
               /clone="U62067"
               /ecotype="Columbia"
               /note="This clone is in pUNI 51"
     CDS   1..2790
               /codon_start=1
               /note="formin homology 2 (FH2) domain-containingprotein"
               /product="At1g59910"
               /protein_id="AAQ62880.1"
               /db_xref="GI:34222088"
               /translation="MSFLFRKNGSSSRRKIKEKLRGRHSDRGGKREEDERYGGSGSDL
               PPPSPWGFLFPBDFVRIDGNLKAVIVDDEGLDVIYWKKLLELENSGKIRKNPKPRRR
               GDKSGDGFRRTGADQDDNDDGDDEVGDESIEEAFSFHVKKSQSASSGGEIRDQSNNG
               GGGGGGGGGGGRYYTSSSASPSRPSSSSASASASPSRTSYATSAGSDYGGGGGKQSQS
               KFQAPGGGSFPSSPSQIHSGGGRSPPLPLPPGQFTAGNASFPSSTQPPPGQYMAGNAS
               FPSSTPPPGQYMAGNAPFSSSTPLPPGQYPAVNAQLSTSAPSVPLPPGQYTAVNAPF
               STSTQPVSLIPPGQYMPGNAALSASTPLITPGQPTTANAPPAPPGPANQTSPPPPPPSA
               AAPPPPPPPKKGPAAPPPPPPPGGKAAGPPPPMSKKGPPFKKGPPKPGNPKGPTKSGETSL
               AVGKTEDPTQPKLKPLHWDKMNPDASRSMVWHKIDGGSFNFDGDLMEALFGYVARKPS
               ESNSVPQNQTVSNSVPHNQTYILDPRKSQNKAIVLKSLGMTKBEIIDLLTEGHDAESD
               TLEKLAGIAPTPEQTEIIDFDGBPMTLAYADSLLFHILKAVPSAFNRFNVMLFKINY
               GSEVAQQKGSLLTLESACNELRARGLFMKLLEAILKAGNRMNAGTARGNAQAFNLTAL
               RKLSDVKSVDAKTLLHFVEVVRSEGKRAAMNKNMMSSDNGSGENADMSREQEIE
               FIKMGLPIIGGLSSEPTNVKKAAGIDYDSFVATTLALGTRVKETKRLLDQSKGKEDGC
               LTKLRSFPBSAEEELKVITEEQLRIMELVKKTNYYQAGALKBRNLFQLFVIRDFLG
               MVDNACSBIARNQRKQQQQRPATTVAGASSSPAETPSVAAAPQRNAVRFPILPNFMS
               ESSRXSSSSDSDSES"

ORIGIN

Query Match   3.0%;   Score 33;   DB 8;   Length 2790;
Best Local Similarity 100.0%; Pred. No. 6.1e-07;
Matches   33;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
            |||||||||||||||||||||||||||||||||||
Db   516 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 484

```
RESULT 13
AC100332
LOCUS       AC100332               59159 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-126G16, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100332
VERSION     AC100332.1  GI:17047698
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 59159)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-126G16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 59159)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassilev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L15119
            Center clone name: 126_G_16
            ---------------
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            *       1      704: contig of 704 bp in length
            *     705      804: gap of 100 bp
            *     805     1534: contig of 730 bp in length
            *    1535     1634: gap of 100 bp
            *    1635     2338: contig of 704 bp in length
            *    2339     2438: gap of 100 bp
            *    2439     3120: contig of 682 bp in length
            *    3121     3220: gap of 100 bp
            *    3221     3907: contig of 687 bp in length
            *    3908     4007: gap of 100 bp
            *    4008     4691: contig of 684 bp in length
            *    4692     4791: gap of 100 bp
            *    4792     5475: contig of 684 bp in length
            *    5476     5575: gap of 100 bp
            *    5576     6258: contig of 683 bp in length
            *    6259     6358: gap of 100 bp
            *    6359     7057: contig of 699 bp in length
            *    7058     7157: gap of 100 bp
            *    7158     7867: contig of 710 bp in length
            *    7868     7967: gap of 100 bp
            *    7968     8647: contig of 680 bp in length
            *    8648     8747: gap of 100 bp
            *    8748     9478: contig of 731 bp in length
            *    9479     9578: gap of 100 bp
            *    9579    10292: contig of 714 bp in length
            *   10293    10392: gap of 100 bp
            *   10393    11088: contig of 696 bp in length
            *   11089    11188: gap of 100 bp
            *   11189    11870: contig of 682 bp in length
            *   11871    11970: gap of 100 bp
            *   11971    12674: contig of 704 bp in length
            *   12675    12774: gap of 100 bp
            *   12775    13485: contig of 711 bp in length
            *   13486    13585: gap of 100 bp
            *   13586    14306: contig of 721 bp in length
            *   14307    14406: gap of 100 bp
            *   14407    15121: contig of 715 bp in length
            *   15122    15221: gap of 100 bp
            *   15222    15953: contig of 732 bp in length
            *   15954    16053: gap of 100 bp
            *   16054    16673: contig of 620 bp in length
            *   16674    16773: gap of 100 bp
            *   16774    17453: contig of 680 bp in length
            *   17454    17553: gap of 100 bp
            *   17554    18238: contig of 685 bp in length
            *   18239    18338: gap of 100 bp
            *   18339    19025: contig of 687 bp in length
            *   19026    19125: gap of 100 bp
            *   19126    19805: contig of 680 bp in length
            *   19806    19905: gap of 100 bp
            *   19906    20604: contig of 699 bp in length
            *   20605    20705: gap of 100 bp
            *   20705    21415: contig of 711 bp in length
            *   21416    21515: gap of 100 bp
            *   21516    22196: contig of 681 bp in length
            *   22197    22296: gap of 100 bp
            *   22297    23023: contig of 727 bp in length
            *   23024    23123: gap of 100 bp
            *   23124    23834: contig of 711 bp in length
            *   23835    23934: gap of 100 bp
            *   23935    24642: contig of 708 bp in length
            *   24643    24742: gap of 100 bp
            *   24743    25428: contig of 686 bp in length
            *   25429    25528: gap of 100 bp
            *   25529    26219: contig of 691 bp in length
            *   26220    26319: gap of 100 bp
            *   26320    27005: contig of 686 bp in length
            *   27006    27105: gap of 100 bp
            *   27106    27813: contig of 708 bp in length
            *   27814    27913: gap of 100 bp
            *   27914    28625: contig of 712 bp in length
            *   28626    28725: gap of 100 bp
            *   28726    29448: contig of 723 bp in length
            *   29449    29548: gap of 100 bp
            *   29549    30271: contig of 723 bp in length
            *   30272    30371: gap of 100 bp
            *   30372    31100: contig of 729 bp in length
            *   31101    31200: gap of 100 bp
            *   31201    31933: contig of 733 bp in length
            *   31934    32033: gap of 100 bp
            *   32034    32721: contig of 688 bp in length
            *   32722    32821: gap of 100 bp
```

```
*  33822    33513: contig of 692 bp in length
*  33514    33613: gap of 100 bp
*  33614    34311: contig of 698 bp in length
*  34412    34411: gap of 100 bp
*  35103    35102: contig of 691 bp in length
*  35203    35202: gap of 100 bp
*  35890    35889: contig of 687 bp in length
*  35990    35988: gap of 100 bp
*  36700    36699: contig of 710 bp in length
*  37508    36799: gap of 100 bp
*  37608    37507: contig of 708 bp in length
*  38286    37607: gap of 100 bp
*  38386    38285: contig of 678 bp in length
*  39116    38385: gap of 100 bp
*  39216    39115: contig of 730 bp in length
*  39937    39215: gap of 100 bp
*  40037    39936: contig of 721 bp in length
*  40749    40036: gap of 100 bp
*  40849    40748: contig of 712 bp in length
*  41540    40848: gap of 100 bp
*  41640    41539: contig of 691 bp in length
*  42330    41639: gap of 100 bp
*  42430    42329: contig of 690 bp in length
*  43121    42429: gap of 100 bp
*  43221    43120: contig of 691 bp in length
*  44019    43220: gap of 100 bp
*  44736    43918: contig of 698 bp in length
*  45545    44018: gap of 100 bp
*  46332    44735: contig of 717 bp in length
*  46432    44835: gap of 100 bp
*  47264    45544: contig of 709 bp in length
*  47988    45644: gap of 100 bp
*  48802    46331: contig of 687 bp in length
*  48902    46431: gap of 100 bp
*  49688    47163: contig of 732 bp in length
*  50379    47263: gap of 100 bp
*  51168    47987: contig of 724 bp in length
*  51268    48087: gap of 100 bp
*  52053    48801: contig of 714 bp in length
*  52746    48901: gap of 100 bp
*  53528    49587: contig of 686 bp in length
*  54341    49687: gap of 100 bp
*  54441    50378: contig of 691 bp in length
*  55143    50478: gap of 100 bp
           51167: contig of 689 bp in length
           51267: gap of 100 bp
           51952: contig of 685 bp in length
           52052: gap of 100 bp
           52745: contig of 693 bp in length
           52845: gap of 100 bp
           53527: contig of 682 bp in length
           53627: gap of 100 bp
           54340: contig of 713 bp in length
           54440: gap of 100 bp
           55142: contig of 702 bp in length
           55242: gap of 100 bp
```

Query Match   3.0%;   Score 33; DB 2; Length 59159;
Best Local Similarity   100.0%; Pred. No. 7.2e-07; Indels 0; Gaps 0;
Matches   33; Conservative 0; Mismatches 0;

```
Qy    973  TCTTCCCTCCTCCTCCTCCTCCTCCTCCACC  1005
           |||||||||||||||||||||||||||||||
Db  50646  TCTTCCCTCCTCCTCCTCCTCCTCCTCCACC  50678
```

RESULT 14
AC140499
LOCUS   AC140499   78244 bp   DNA   linear   HTG 26-MAY-2003
DEFINITION Mus musculus chromosome 8 clone RP23-439N24 map 8, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC140499
VERSION   AC140499.2 GI:31075328
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 78244)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Mus musculus chromosome 8, clone RP23-439N24
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 78244)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 78244)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 26, 2003 this sequence version replaced gi:28475865.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
            Center project name: L28787
            Center clone name: 439_N_24
            ----
            * NOTE: This record contains 76 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads

```
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        934: contig of 934 bp in length
*   935 1034: gap of 100 bp
*  1035 1964: contig of 930 bp in length
*  1965 2064: gap of 100 bp
*  2065 2968: contig of 904 bp in length
*  2969 3068: gap of 100 bp
*  3069 3975: contig of 907 bp in length
*  3976 4075: gap of 100 bp
*  4076 4997: contig of 922 bp in length
*  4998 5097: gap of 100 bp
*  5098 6020: contig of 923 bp in length
*  6021 6120: gap of 100 bp
*  6121 7051: contig of 931 bp in length
*  7052 7151: gap of 100 bp
*  7152 8081: contig of 930 bp in length
*  8082 8181: gap of 100 bp
*  8182 9132: contig of 951 bp in length
*  9133 9232: gap of 100 bp
*  9233 10173: contig of 941 bp in length
* 10174 10273: gap of 100 bp
* 10274 11203: contig of 930 bp in length
* 11204 11303: gap of 100 bp
* 11304 12224: contig of 921 bp in length
* 12225 12324: gap of 100 bp
* 12325 13200: contig of 876 bp in length
* 13201 13300: gap of 100 bp
* 13301 14247: contig of 947 bp in length
* 14248 14347: gap of 100 bp
* 14348 15310: contig of 963 bp in length
* 15311 15410: gap of 100 bp
* 15411 16348: contig of 938 bp in length
* 16349 16448: gap of 100 bp
* 16449 17390: contig of 942 bp in length
* 17391 17490: gap of 100 bp
* 17491 18432: contig of 942 bp in length
* 18433 18532: gap of 100 bp
* 18533 19470: contig of 938 bp in length
* 19471 19570: gap of 100 bp
* 19571 20488: contig of 918 bp in length
* 20489 20588: gap of 100 bp
* 20589 21527: contig of 939 bp in length
* 21528 21627: gap of 100 bp
* 21628 22541: contig of 914 bp in length
* 22542 22641: gap of 100 bp
* 22642 23583: contig of 942 bp in length
* 23584 23683: gap of 100 bp
* 23684 24594: contig of 911 bp in length
* 24595 24694: gap of 100 bp
* 24695 25629: contig of 935 bp in length
* 25630 25729: gap of 100 bp
* 25730 26667: contig of 938 bp in length
* 26668 26767: gap of 100 bp
* 26767 27714: contig of 947 bp in length
* 27715 27814: gap of 100 bp
* 27815 28738: contig of 924 bp in length
* 28739 28838: gap of 100 bp
* 28839 29769: contig of 931 bp in length
* 29770 29869: gap of 100 bp
* 29870 30781: contig of 912 bp in length
* 30782 30881: gap of 100 bp
* 30881 31803: contig of 922 bp in length
* 31804 31903: gap of 100 bp
* 31904 32828: contig of 925 bp in length
* 32829 32928: gap of 100 bp
* 32929 33891: contig of 963 bp in length
* 33892 33991: gap of 100 bp
* 33992 34957: contig of 966 bp in length
* 34958 35057: gap of 100 bp
* 35058 35997: contig of 940 bp in length
* 35998 36097: gap of 100 bp
* 36098 37035: contig of 938 bp in length
* 37036 37135: gap of 100 bp
* 37136 38088: contig of 953 bp in length
* 38089 38188: gap of 100 bp
* 38189 39119: contig of 931 bp in length
* 39120 39219: gap of 100 bp
* 39220 40142: contig of 923 bp in length
* 40143 40242: gap of 100 bp
* 40243 41157: contig of 915 bp in length
* 41158 41257: gap of 100 bp
* 41258 42196: contig of 939 bp in length
* 42197 42296: gap of 100 bp
* 42297 43225: contig of 929 bp in length
* 43226 43325: gap of 100 bp
* 43326 44290: contig of 965 bp in length
* 44291 44390: gap of 100 bp
* 44391 45344: contig of 954 bp in length
* 45345 45444: gap of 100 bp
* 45445 46388: contig of 944 bp in length
* 46389 46488: gap of 100 bp
* 46489 47415: contig of 927 bp in length
* 47416 47515: gap of 100 bp
* 47516 48455: contig of 940 bp in length
* 48456 48555: gap of 100 bp
* 48556 49498: contig of 943 bp in length
* 49499 49598: gap of 100 bp
* 49599 50527: contig of 929 bp in length
* 50528 50627: gap of 100 bp
* 50628 51547: contig of 920 bp in length
* 51548 51647: gap of 100 bp
* 51648 52587: contig of 940 bp in length
* 52588 52687: gap of 100 bp
* 52688 53660: contig of 973 bp in length
* 53661 53760: gap of 100 bp
* 53761 54711: contig of 951 bp in length
* 54712 54811: gap of 100 bp
* 54812 55753: contig of 942 bp in length
* 55754 55853: gap of 100 bp
* 55854 56764: contig of 911 bp in length
* 56765 56864: gap of 100 bp
* 56865 57819: contig of 955 bp in length
* 57820 57919: gap of 100 bp
* 57920 58849: contig of 930 bp in length
* 58850 58949: gap of 100 bp
```

```
Query Match     3.0%;    Score 33;  DB 2; Length 78244;
Best Local Similarity 100.0%;  Pred. No. 7.3e-07;
Matches  33;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          |||||||||||||||||||||||||||||||||
Db  74932 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 74964

RESULT 15
AP004394
LOCUS       AP004394              98286 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 2 clone
            OJ1118_C06, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP004394
VERSION     AP004394.1  GI:17402600
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
```

AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE   Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
          clone:OJ1118_C06

JOURNAL   Published Only in Database (2001)
REFERENCE 2 (bases 1 to 98286)
AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE   Direct Submission
JOURNAL   Submitted (05-DEC-2001) Takuji Sasaki, National Institute of
          Agrobiological Sciences, Rice Genome Research Program; Kannondai
          2-1-2, Tsukuba, Ibaraki 305-8602, Japan
          (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
          Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT   The nucleotide sequence of this BAC clone was generated by
          combining Monsanto and RGP-Japan sequencing data.
          NOTE: It currently consists of 1 contigs. Gaps between the contigs
          are represented as runs of N. The order of the pieces is believed
          to be correct as given, however the sizes of the gaps between them
          are based on estimates that have provided by the submitter. This
          sequence will be replaced by the finished sequence as soon as it is
          available and the accession number will be preserved.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.

FEATURES   Location/Qualifiers
     source   1..98286
                   /organism="Oryza sativa (japonica cultivar-group)"
                   /mol_type="genomic DNA"
                   /cultivar="Nipponbare"
                   /db_xref="taxon:39947"
                   /chromosome="2"
                   /clone="OJ1118_C06"

ORIGIN

Query Match   3.0%; Score 33; DB 2; Length 98286;
Best Local Similarity   100.0%; Pred. No. 7.4e-07;
Matches 33; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
           ||||||||||||||||||||||||||||||||||||
Db   41162 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCCACC 41194

Search completed: February 21, 2004, 22:57:51
Job time : 4364 secs

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      February 21, 2004, 19:27:14 ; Search time 483 Seconds
                                  (without alignments)
                              9710.167 Million cell updates/sec

Title:       US-10-041-615-33
Perfect score: 1104
Sequence:    1 atgataaaactaggcccta.........gtgtcattgctttatgact 1104
Scoring table: OLIGO_NUC
             Gapop 60.0 , Gapext 60.0
Searched:    3373863 seqs, 2124099041 residues
Word size :  10

Total number of hits satisfying chosen parameters:      1958023

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
             1: geneseqn1980s:*
             2: geneseqn1990s:*
             3: geneseqn2000s:*
             4: geneseqn2001as:*
             5: geneseqn2001bs:*
             6: geneseqn2002s:*
             7: geneseqn2003as:*
             8: geneseqn2003bs:*
             9: geneseqn2003cs:*
            10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1104	100.0	1104	6	ABS64723	Abs64723 cDNA enco
2	461	41.8	4319	9	ADC86840	Adc86840 Human GPC
3	461	41.8	9409	9	AAK76477	Aak76477 Human imm
c 4	458	41.5	1001	6	AAS98131	Aas98131 Human DNA
5	409	37.0	1526	6	ABK94922	Abk94922 Human nov
6	318	28.8	330	4	AAS98111	Aas98111 Human DNA
7	279	25.3	1771	5	AAS78869	Aas78869 DNA encod
8	176	15.9	536	4	AAS42840	Aas42840 Human G P
c 9	159	14.4	556	4	AAS30791	Aas30791 Human cDN
c 10	103	9.3	594	6	AAS98061	Aas98061 Human DNA
11	33	3.0	22760	4	ABA15489	Aba15489 Human ner
12	31	2.8	232	4	AAI22451	Aai22451 Probe #12
13	31	2.8	232	4	ABA67527	Aba67527 Human foe
14	31	2.8	232	4	AAI47743	Aai47743 Probe #16
15	31	2.8	232	4	ABA49623	Aba49623 Human bre
16	31	2.8	232	4	ABA34608	Aba34608 Probe #13
17	31	2.8	232	4	AAK41698	Aak41698 Human bon
18	31	2.8	232	4	AAK15953	Aak15953 Human liv
19	31	2.8	232	4	ABS41287	Abs41287 Human liv
20	31	2.8	232	5	AAI08137	Aai08137 Probe #81
21	31	2.8	232	6	ABS15705	Abs15705 Human gen
22	31	2.8	384	4	AAI13216	Aai13216 Probe #31
23	31	2.8	384	4	ABA54915	Aba54915 Human foe
24	31	2.8	384	4	AAI34569	Aai34569 Probe #32
25	31	2.8	384	4	ABA44472	Aba44472 Human bre
26	31	2.8	384	4	ABA24683	Aba24683 Probe #31
27	31	2.8	384	4	AAK28643	Aak28643 Human bon
28	31	2.8	384	4	AAK03190	Aak03190 Human bra
29	31	2.8	384	4	ABS28239	Abs28239 Human liv
30	31	2.8	384	5	AAI03121	Aai03121 Probe #31
31	31	2.8	384	5	ABS03152	Abs03152 Human gen
c 32	31	2.8	1728	3	AAA92174	Aaa92174 Oryzias l
c 33	31	2.8	1863	3	AAA92175	Aaa92175 Oryzias l
c 34	31	2.8	2764	3	AAD31046	Aad31046 Medaka fi
c 35	31	2.8	6115	3	AAZ46519	Aaz46519 Murine CA
c 36	30	2.7	69	3	AAC11384	Aac11384 Human sec
37	30	2.7	123	4	ABA70569	Aba70569 Human foe
38	30	2.7	123	4	AAI50733	Aai50733 Probe #19
39	30	2.7	123	4	AAK44756	Aak44756 Human bon
40	30	2.7	123	4	AAK18816	Aak18816 Human bra
41	30	2.7	123	4	ABS44416	Abs44416 Human liv
42	30	2.7	123	6	ABS18995	Abs18995 Human gen
43	30	2.7	135	5	AAS80508	Aas80508 DNA encod
44	30	2.7	207	4	AAI24964	Aai24964 Probe #14
45	30	2.7	207	4	ABA70492	Aba70492 Human foe

```
                             ALIGNMENTS

RESULT 1
ABS64723
ID   ABS64723 standard; cDNA; 1104 BP.
XX
AC   ABS64723;
XX
DT   15-NOV-2002  (first entry)
XX
DE   cDNA encoding human GPCR16 protein.
XX
KW   G protein coupled receptor related protein; ss; gene; human; GPCR;
KW   cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
KW   Von Hippel-Lindau syndrome; Alzheimer's disease; tuberous sclerosis;
KW   hypercalcaemia; Parkinson's disease; Huntingtons's disease;
KW   cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW   ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW   pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW   inflammatory bowel disease; infertility; hypertension; scleroderma;
KW   arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW   infection; graft-versus-host disease.
XX
OS   Homo sapiens.
XX
PN   WO20264793-A2.
XX
PD   22-AUG-2002.
XX
PF   03-JAN-2002; 2002WO-US000056.
XX
PR   03-JAN-2001; 2001US-0259552P.
PR   09-JAN-2001; 2001US-0260544P.
PR   20-MAR-2001; 2001US-0274405P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Casman SJ, Edinger SR, Ellerman K, Smithson G, Kekuda R;
PI   Muralidhara P;
XX
DR   WPI; 2002-643487/69.
DR   P-PSDB; ABG32286.
XX
PT   New isolated G protein coupled receptor polypeptides and polynucleotides,
PT   useful in gene therapy, particularly for treating or preventing
PT   cardiomyopathy, atherosclerosis, diabetes, multiple sclerosis, Crohn's
PT   disease or cancer.
XX
```

```
PS   Claim 9; Page 97; 252pp; English.
XX
CC   This invention relates to a new isolated G-protein coupled receptor
CC   (GPCRX) polypeptide sequence and the cDNA encoding it. The GPCRX
CC   polypeptide. GPCRX nucleic acid and an antibody specific to the protein
CC   are useful for treating, preventing or alleviating a GPCRX-associated
CC   disorder or a pathological state in a subject, particularly a human. In
CC   particular, the disorder is cardiomyopathy, atherosclerosis, diabetes, or
CC   a disorder related to cell signal processing and metabolic pathway
CC   modulation. The GPCRX polypeptide and nucleic acid are also useful for
CC   diagnosing the presence of or predisposition to a disease associated with
CC   altered levels of GPCRX, particularly cancer. The GPCRX nucleic acid and
CC   polypeptide are especially useful in the manufacture of a medicament for
CC   therapeutic or prophylactic applications for disorders associated with
CC   aberrant GPCRX expression or activity, e.g. Von Hippel-Lindau syndrome,
CC   Alzheimer's disease, stroke, tuberous sclerosis, hypercalcaemia,
CC   Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy,
CC   Lesch-Nyhan syndrome, multiple sclerosis, ataxia-telangiectasia,
CC   leukodystrophies, addiction, anxiety, depression, pain, obesity, Crohn's
CC   disease, osteoporosis, inflammatory bowel disease, infertility,
CC   hypertension, scleroderma, haemophilia, asthma, arthritis, human
CC   immunodeficiency virus; autoimmune disease; HIV, viral, fungal, bacterial
CC   or protozoal infections, or graft-versus -host disease. The DNA encoding
CC   the protein is useful in gene therapy for treating the above conditions.
CC   The polypeptides can be used as immunogens to produce antibodies and as
CC   vaccines. The nucleic acids are further used as hybridisation probes, in
CC   chromosome mapping, tissue typing, preventive medicine, and
CC   pharmacogenomics. These are also useful in developing powerful assay
CC   system for functional analysis of various human disorders, as well as in
CC   diagnostic applications. The present sequence represents the cDNA
CC   encoding a human G protein coupled receptor related protein (GPCR) of the
CC   invention
XX
SQ   Sequence 1104 BP; 217 A; 328 C; 290 G; 269 T; 0 U; 0 Other;

Query Match       100.0%;  Score 1104;  DB 6;  Length 1104;
Best Local Similarity  100.0%;  Pred. No.0;
Matches 1104;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 ATGATAAAACTAGGCCCTTACTTCACACCACCCACAAAAAAAATAAAACTAAGATTAAAGAT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGATAAAACTAGGCCCTTACTTCACACCACCCACAAAAAAAATAAAACTAAGATTAAAGAT 60

Qy   61 TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGCGCTCTGTTCTACACAGCCCTC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTTGCGCTCTGTTCTACACAGCCCTC 120

Qy  121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTTACCTGTCAAAAGAGCAGG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTTACCTGTCAAAAGAGCAGG 180

Qy  181 AAGATCAACTGCACAGGCATCACCTACCTGGTCGCACCTGGCTGTCTGCTCGTGTTCACC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGATCAACTGCACAGGCATCACCTACCTGGTCGCACCTGGCTGTCTGCTCGTGTTCACC 240

Qy  241 GTGGCCTTACCGGGGAAGGGGTGGTGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGCAAG 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GTGGCCTTACCGGGGAAGGGGTGGTGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGCAAG 300

Qy  301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACACACTTACGGGGGGGTCTACCTC 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACACACTTACGGGGGGGTCTACCTC 360

Qy  361 ATGGCCTGTGTGAGCGGTGAGCGGTGACCATTACCCAGCTGTGGTCTGTGCCGGCCCGCGC 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 ATGGCCTGTGTGAGCGGTGAGCGGTGACCATTACCCAGCTGTGGTCTGTGCCGGCCCGCGC 420

Qy  421 CTCCGCACGGCTGCCGCGCCGCCAGGCTGCTTGCGGCCATCTGGACCTTGGCTGTGCTG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 CTCCGCACGGCTGCCGCGCCGCCAGGCTGCTTGCGGCCATCTGGACCTTGGCTGTGCTG 480

Qy  481 CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGTGGTGGGCAAGCTGGCCTGC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGTGGTGGGCAAGCTGGCCTGC 540

Qy  541 ATGGAGTACAGCAGCAGTGAGTCAGTCTCGGGTGCCCTCATGGTCCTGGTGCCCTTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 ATGGAGTACAGCAGCAGTGAGTCAGTCTCGGGTGCCCTCATGGTCCTGGTGCCCTTT 600

Qy  601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCGTCCTGCTATATGAAGATCACCTGG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 GCCATTGGCTTCTGTGGGCCAGTGGGGATCATCGTCCTGCTATATGAAGATCACCTG 660

Qy  661 AAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGACCAGCAGGAAAGGACGACCCACTGGCGA 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 AAGCTGTGCAGCACAGCTCGGGAGGACCCAGTGACCAGCAGGAAAGGACGACCCACTGGCGA 720

Qy  721 GGCTGCCTGCTTACCGCTGCTGATGCTGGTGGCCGTGGTGCTCGTTCAGCCCCTACCAC 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GGCTGCCTGCTTACCGCTGCTGATGCTGGTGGCCGTGGTGCTCGTTCAGCCCCTACCAC 780

Qy  781 CTCAACATCAAGCAGTTCATGGCGAGAGGATGTCCACCTGCCATCCTGCCGAGCGG 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CTCAACATCAAGCAGTTCATGGCGAGAGGATGTCCACCTGCCATCCTGCCGAGCGG 840

Qy  841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCTGGCCCATCATGAACATGAACTGTGGCATT 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCTGGCCCATCATGAACATGAACTGTGGCATT 900

Qy  901 ACCCCAATCATTACTTCTTTGCATCCACCCATTACAGGAAATGGCTCCTGGGCATTTTA 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 ACCCCAATCATTACTTCTTTGCATCCACCCATTACAGGAAATGGCTCCTGGGCATTTTA 960

Qy  961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCAGGAAAAGCTTCT 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCAGGAAAAGCTTCT 1020

Qy 1021 TCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCTTATGTTCTTAGCGGAGCACGTGGTC 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 TCAGAAACACCAAGTATCACCCAGGCCAGAGGTTCTTATGTTCTTAGCGGAGCACGTGGTC 1080

Qy 1081 TAACGTGTCATTTGCTTTATGACT 1104
         ||||||||||||||||||||||||
Db 1081 TAACGTGTCATTTGCTTTATGACT 1104
```

```
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1293; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 4319 BP; 1058 A; 1049 C; 1170 G; 994 T; 0 U; 48 Other;

Query Match             41.8%;  Score 461;  DB 9;  Length 4319;
Best Local Similarity   99.8%;  Pred. No. 9.4e-217;
Matches  581;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy      97 CGTCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGAAACATCCTTGCCCTT    156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2940 CGTCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGAAACATCCTTGCCCTT   2999

Qy     157 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGTGCACCTG    216
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3000 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGTGCACCTG   3059

Qy     217 GCTGTGTCTGACCTGCTGTTCACCGGTGCCTTACCGGGAAGGTGGTGTGTTATGTGCTG    276
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3060 GCTGTGTCTGACCTGCTGTTCACCGGTGCCTTACCGGGAAGGTGGTGTGTTATGTGCTG   3119

Qy     277 GGCTCCAGCTGGCCTTTCGGCAAGGGCTCTGCAGGCTGACGGCGTTGTGTGCTCTACACC    336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3120 GGCTCCAGCTGGCCTTTCGGCAAGGGCTCTGCAGGCTGACGGCGTTGTGTGCTCACACC   3179

Qy     337 GACACCTACGGGGGGGGTCTACCTCATGGCCTGTGTGAGCGTGAGCGTGGACCATTACCCAGCTGTG    396
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    3180 GACACCTAC-GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGAGCGTGGACCATTACCCAGCTGTG   3238

Qy     397 GTCTGTCCCACTGGGGCCCGGCCCTCCGCACGGCTGGCGCCCAGGCTGGTCTGCGTG    456
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3239 GTCTGTGCCCACTGGGGCCCGGCCCTCCGCACGGCTGGCGCCCAGGCTGGTCTGCGTG   3298

Qy     457 GCCATCTCGACCTTGGTGCTGCGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG    516
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3299 GCCATCTGGACCTTGGTGCTGCGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG   3358

Qy     517 CCGCTGGTGGGCAAGCTGGCCTTGCATGGAGTCAGCAGCATGGAGTCAGTCCTCGGGCTG    576
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3359 CCGCTGGTGGGCAAGCTGGCCTTGCATGGAGTCAGCAGCATGGAGTCAGTCCTCGGGCTG   3418

Qy     577 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTG    636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3419 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTG   3478

Qy     637 TCCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT    678
           ||||||||||||||||||||||||||||||||||||||||||
Db    3479 TCCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT   3520
```

```
RESULT 3
AAK76477
ID   AAK76477 standard; DNA; 9409 BP.
XX
AC   AAK76477;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31289.
XX.
```

```
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   24-FEB-2000;  2000US-0184464P.
PR   02-MAR-2000;  2000US-0186350P.
PR   16-MAR-2000;  2000US-0189874P.
PR   17-MAR-2000;  2000US-0190076P.
PR   18-APR-2000;  2000US-0198123P.
PR   19-MAY-2000;  2000US-0205515P.
PR   07-JUN-2000;  2000US-0209467P.
PR   28-JUN-2000;  2000US-0214886P.
PR   30-JUN-2000;  2000US-0215135P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0209963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225213P.
PR   14-AUG-2000;  2000US-0225214P.
PR   14-AUG-2000;  2000US-0225266P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0225447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   14-AUG-2000;  2000US-0225759P.
PR   18-AUG-2000;  2000US-0226279P.
PR   22-AUG-2000;  2000US-0226681P.
PR   22-AUG-2000;  2000US-0226868P.
PR   22-AUG-2000;  2000US-0227182P.
PR   23-AUG-2000;  2000US-0227009P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   06-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   08-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   08-SEP-2000;  2000US-0232081P.
PR   12-SEP-2000;  2000US-0231968P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
PR   14-SEP-2000;  2000US-0232399P.
PR   14-SEP-2000;  2000US-0232400P.
PR   14-SEP-2000;  2000US-0232401P.
PR   14-SEP-2000;  2000US-0233063P.
PR   14-SEP-2000;  2000US-0233064P.
PR   14-SEP-2000;  2000US-0233065P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
```

PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 31289; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 9409 BP; 2123 A; 2163 C; 2259 G; 2864 T; 0 U; 0 Other;

Query Match   41.8%; Score 461; DB 4; Length 9409;
Best Local Similarity 99.8%; Pred. No. 9.4e-217;
Matches 581; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy   97 CGCTCTCGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 CGCTCTCGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 432

Qy   157 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 216
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 492

Qy   217 GCTGTGTCGACCTGGTTCACCGTGGCCTTACCGGGGAGGGTGGTGTGTTATGTGCTG 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 GCTGTGTCGACCTGGTTCACCGTGGCCTTACCGGGGAGGGTGGTGTGTTATGTGCTG 552

Qy   277 GGCTCCAGCTGGCCTTGGGAAGGGGCTCTGCAGGGCTGACGGCGTTGTGCTCTACACC 336
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 GGCTCCAGCTGGCCTTGGGAAGGGGCTCTGCAGGGCTGACGGCGTTGTGCTCTACACC 612

Qy   337 GACACCTACGGGGGGCTCTACCTCATGGCCTGTGTGGAGCGTGGACCATTACCCAGCTGTG 396
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   613 GACACCTAC-GGGGGGCTCTACCTCATGGCCTGTGTGGAGCGTGGACCATTACCCAGCTGTG 671

Qy   397 GTCTGTGCCCACTGGGGCCCGGGCCTCCGCACGGCGTGGCCCGCGCTGGTCTCGCGTG 456
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   672 GTCTGTGCCCACTGGGGCCCGGGCCTCCGCACGGCGTGGCCCGCGCTGGTCTCGCGTG 731

Qy   457 GCCATCTGGACCTTGGTGCGTGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG 516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   732 GCCATCTGGACCTTGGTGCGTGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG 791

Qy   517 CGCTGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTG 576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   792 CGCTGGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGGCTG 851

Qy   577 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGTTCGTGGGGCCAGTGGGGATCATCCTG 636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   852 CCCCTCATGGTCCTGGTGGCCTTTGCCATTGGTTCGTGGGGCCAGTGGGGGATCATCCTG 911

PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

```
Qy    637  TCCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT  678
           ||||||||||||||||||||||||||||||||||||||||
Db    912  TCCTGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT  953
```

RESULT 4
AAS98131/c
ID   AAS98131 standard; cDNA; 1001 BP.
XX
AC   AAS98131;
XX
DT   12-MAR-2002 (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #88.
XX
KW   Human; G protein-coupled receptor; GPCR; ss; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity; diabetes;
KW   hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US015332.
XX
PR   11-MAY-2000; 2000US-0203217P.
PR   18-MAY-2000; 2000US-0205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin receptor
PT   polypeptides useful for identifying modulators that are useful for
PT   treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT   stroke.
XX
PS   Claim 2; Page 95; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a nucleic
CC   acid molecule that is at least 80% identical to the G protein-coupled
CC   (GPCR) polynucleotides included in the specification. Also included are
CC   probes based on the GPCR sequences (including antisense probes), a host
CC   cell comprising an expression vector comprising the GPCR sequence,
CC   antibodies raised against the polypeptides, and methods of identifying
CC   modulators of the polypeptides. The polypeptides are useful for
CC   identifying modulator compounds which function as modulators, activators,
CC   repressors, agonists or antagonists of the novel GPCR polypeptides
CC   including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC   described above can be used to detect the presence of the polypeptides
CC   and nucleic acids and are used to diagnose a variety of diseases or
CC   disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC   amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC   carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC   obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC   macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC   osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC   arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC   diseases listed in the specification. The probes and antibodies are also
CC   useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC   release disorders, cardiovascular activity disorders, pain perception
```

```
CC   disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC   that decrease or increase the expression of galanin receptor (GAL4) can
CC   be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC   nucleic acid is useful for treating the above mentioned disorders by gene
CC   therapy techniques. The present sequence is a novel GPCR polynucleotide
CC   of the invention
XX
SQ   Sequence 1001 BP; 190 A; 315 C; 311 G; 185 T; 0 U; 0 Other;

Query Match      41.5%;  Score 458;  DB 6;  Length 1001;
Best Local Similarity  99.8%;  Pred. No. 2.8e-215;
Matches  578;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;
```

```
Qy    100  TCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGC  159
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1001  TCTCTGTTCTACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGC  942

Qy    160  CTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCT  219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    941  CTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGGCT  882

Qy    220  GTGTCTGACCTGCTGTTCACCGTGCCTTACCGTGGCGAAGGGTGTGTTATGTGCTGGGGC  279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    881  GTGTCTGACCTGCTGTTCACCGTGCCTTACCGTGGCGAAGGGTGTGTTATGTGCTGGGGC  822

Qy    280  TCCAGTGGCCTTTCGGCAAGGGCTCTGCAGGCGTTCGCAGGCTGACGGGCGTTTGTCCTTCCTCACACCGAC  339
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    821  TCCAGTGGCCTTTCGGCAAGGGCTCTGCAGGCGTTCGCAGGCTGACGGGCGTTTGTCCTTCCTCACACCGAC  762

Qy    340  ACCTACGGGGGGGTCTACCTCATGGGCCGGCTCCGCACGGCTGGCGCCGCCAGGCTGGTCTGCGTGGCC  399
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    761  ACCTAC-GGGGGGTCTACCTCATGGGCCGGCTCCGCACGGCTGGCGCCGCCAGGCTGGTCTGCGTGGCC  703

Qy    400  TGTGCCCACTGGGGCCCCGGCGCCTCCGCAGAGATGGCCCTTGCTCTTGTCTTGCTCTTGATGCCCATGACCAAGCCG  459
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    702  TGTGCCCACTGGGGCCCCGGCGCCTCCGCAGAGATGGCCCTTGCTCTTGTCTTGCTCTTGATGCCCATGACCAAGCCG  643

Qy    460  ATCTGGACCTTGGTGCTGCTGCAGAGATGGCCCTTGCTCTTGTCTTGCTCTTGATGCCCATGACCAAGCCG  519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    642  ATCTGGACCTTGGTGCTGCTGCAGAGATGGCCCTTGCTCTTGTCTTGCTCTTGATGCCCATGACCAAGCCG  583

Qy    520  CTGGTGGGCAAGCTGGCCTGCCTGCATGAGTACAGCAGCATGGGCTCAGTCCTCCGGGGCTGCCC  579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    582  CTGGTGGGCAAGCTGGCCTGCCTGCATGAGTACAGCAGCATGGGCTCAGTCCTCCGGGGCTGCCC  523

Qy    580  CTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGCGATCATCCGTCC  639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    522  CTCATGGTCCTGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGCGATCATCCGTCC  463

Qy    640  TGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT  678
           ||||||||||||||||||||||||||||||||||||||
Db    462  TGCTATATGAAGATCACCTGAAGCTGTGCAGCACAGCT  424
```

RESULT 5
ABK94922
ID   ABK94922 standard; cDNA; 1526 BP.
XX
AC   ABK94922;
XX
DT   30-AUG-2002 (first entry)
XX
DE   Human novel polynucleotide #33.
XX
KW   Human; gene; ss; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombosis; coagulation disorder;
KW   fungal infection.

```
Db   662  ACACCTAC-GGGGGGTCTACCTCATGGCTCATGTGTGAGGCGTGGACCATTACCCAGCTGTGG 720
               ||||||||||||  |||| || ||||||||  |  | |||| ||||| ||||||  ||||||
Qy   398  TCTGTGCCACTGGGGGCCCGGCCCTCCGCACGGCTGGCCTCCGCACGGCTGGTCTGCGTGG 457

Db   721  TCTGTGCCACTGGGGGCCCGGCCCTGCTCCGCACGGCTGGCCCAGGCTGGTCTGCGTGG 780
               |||||||||||||||||||||||  ||||||||||||||||| |||||||||||||||||
Qy   458  CCATCTGACCTTGGTGTGCTGCAGAGACGATGCCTTGCTCTTGATGCCATGACCAAGC 517

Db   781  CCATCTGACCTTGGTGTGCTGCAGAGACGATGCCCTTGCTCTTGATGCCATGACCAAGC 840
               ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Qy   518  CGCTGGTGGGCAAGCTGCCTGCCTGCATGAGTACAGCAGCATGAGTCAGTCCTCGGGCTGC 577

Db   841  CGCTGGTGGGCAAGCTGCCTGCCTGCATGAGTACAGCAGCATGAGTCAGTCCTCGGGCTGC 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   578  CCCTCATGTCTCCGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGT 637

Db   901  CCCTCATGTCTCCGGTGGCCTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGT 960
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   638  CCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT 678
               |||||||||||||||||||||||||||||||||||||||||
Db   961  CCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT 1001
```

RESULT 6
AAS98111
ID   AAS98111 standard; DNA; 330 BP.
XX
AC   AAS98111;
XX
DT   12-MAR-2002 (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #68.
XX
KW   Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity; diabetes;
KW   hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US015332.
XX
PR   11-MAY-2000; 2000US-0203217P.
PR   18-MAY-2000; 2000US-0205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin receptor
PT   polypeptides useful for identifying modulators that are useful for
PT   treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT   stroke.
XX
PS   Claim 2; Page 83; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a nucleic
CC   acid molecule that is at least 80% identical to the G protein-coupled
CC   (GPCR) polynucleotides included in the specification. Also included are
CC   probes based on the GPCR sequences (including antisense probes), a host

XX
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US047004.
XX
PR   30-NOV-2000; 2000US-00728952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI   Yamazaki V, Ujwal ML, Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   P-PSDB; ABG66698.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders, cancer and promoting wound healing.
XX
PS   Claim 1; Page 403-404; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue,
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABK94890-ABK94982 represent human
CC   novel polynucleotides of the invention
XX
SQ   Sequence 1526 BP; 282 A; 458 C; 458 G; 328 T; 0 U; 0 Other;

Query Match   37.0%; Score 409; DB 6; Length 1526;
Best Local Similarity   99.7%; Pred. No. 4.1e-191;
Matches 579; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy    98  GCTCTCTGTTCTCACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422  GCTCTCTGTTCTCACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTT 481

Qy   158  GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGG 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482  GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTGG 541

Qy   218  CTGTGCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGGTGGTGTTATGTGCTGG 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542  CTGTGCTGACCTGCTGTTCACCGTGGCCTTACCGGGAAGGGGTGGTGTTATGTGCTGG 601

Qy   278  GCTCCAGCTGGCCTTTCCGGCAAGGGGCTCTGCAGGCTGACGCGGCGTTTGTGCTCTCACACCG 337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602  GCTCCAGCTGGCCTTTCCGGCAAGGGGCTCTGCAGGCTGACGCGGCGTTTGTGCTCTCACACCG 661

Qy   338  ACACCTACGGGGGGGTCTACCTCATGGCTCATGTGTGAGGCGTGGACCATTACCCAGCTGTGG 397
              |||||| ||||||  ||||||||||||||||||||||||||||||||||||||||||||||
```

PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG14682.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 14673; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences.(I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes.(I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1771 BP; 313 A; 531 C; 519 G; 407 T; 0 U; 1 Other;

Query Match   25.3%; Score 279; DB 5; Length 1771;
Best Local Similarity   99.4%; Pred. No. 5.4e-127;
Matches 499; Conservative   0; Mismatches   2; Indels   1; Gaps   1;

Qy   178 AGGAAGATCAACTGCACAGGCATCACCTGGTGCACCTGGCTGTGTCTGGACCTGCTGTTC 237
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   742 AGGAAGATCAACTGCACAGGCATCACCTGGTGCACCTGGCTGTGTCTGGACCTGCTGTTC 801

Qy   238 ACCGTGGCCTTACCGGGAAGGTGGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGC 297
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   802 ACCGTGGCCTTACCGGGAAGGTGGTGTTATGTGCTGGGCTCCAGCTGGCCTTTCGGC 861

Qy   298 AAGGGGCTCTGCAGGCTGACGGCCGTTTGTGCTCTCACACCGACACCTACGGGGGGGTCTAC 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862 AAGGGGCTCTGCAGGCTGACGGCCGTTTGTGCTCTCACACCGACACCTACGGGGGGGTCTAC 921

Qy   358 CTTCATGGCCTGTGTGAGCGTGGACCATTACCCAGC-TGTGGTCGTGTCGTGACCTGCTGTTC 416
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   922 CTTCATGGCCTGTGTGAGCGTGGACCATTACCCAACTTGTGTCGTGTCGTGACCTGCTGTTC 981

Qy   417 GCGGCTTCCGCACGGCTGGCCGCCCAGGCGGTCGGTGGCCATCTGGACCTTCGGCTGCT 476
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982 GTGGCTTCCGCACGGCTGGCCGCCCAGGCGGTCGGTGGCCATCTGGACCTTCGGCTGCT 1041

Qy   477 GCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCGCTCGGTGGGCAAGCTGGC 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 GCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTCGGTGGGCAAGCTGGC 1101

CC   cell comprising an expression vector comprising the GPCR sequence,
CC   antibodies raised against the polypeptides, and methods of identifying
CC   modulators of the polypeptides. The polypeptides are useful for
CC   identifying modulator compounds which function as modulators, activators,
CC   repressors, agonists or antagonists of the novel GPCR polypeptides
CC   including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC   described above can be used to detect the presence of the polypeptides
CC   and nucleic acids and are used to diagnose a variety of diseases or
CC   disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC   amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC   carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC   obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC   macular degeneration lymphoma, melanoma, multiple sclerosis,
CC   osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC   arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC   diseases listed in the specification. The probes and antibodies are also
CC   useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC   release disorders, cardiovascular activity disorders, pain perception
CC   disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC   that decrease or increase the expression of galanin receptor (GAL4) can
CC   be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC   nucleic acid is useful for treating the above mentioned disorders by gene
CC   therapy techniques. The present sequence is a novel GPCR polynucleotide
CC   of the invention
XX
SQ   Sequence 330 BP; 46 A; 101 C; 108 G; 75 T; 0 U; 0 Other;

Query Match   28.8%; Score 318; DB 6; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.1e-146;
Matches 318; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   349 GGGGTCACCTCATGGCCTGCTGTGAGCGTGGACCATTACCCAGCTGTGGTCGTGTGCCCAC 408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGGTCACCTCATGGCCTGCTGTGAGCGTGGACCATTACCCAGCTGTGGTCGTGTGCCCAC 60

Qy   409 TGGGGCCCGCGCCTCCGCACGGCTGGCCGCCCAGGCTGGTCTGCGTGGCCATCTGGACC 468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGGGGCCCGCGCCTCCGCACGGCTGGCCGCCCAGGCTGGTCTGCGTGGCCATCTGGACC 120

Qy   469 TTGGTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGC 528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TTGGTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGC 180

Qy   529 AAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCCGGGCTGCCCCTCATGGTC 588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCCGGGCTGCCCCTCATGGTC 240

Qy   589 CTGGTGGCCTTTGCCATGGCTTCTGTGGGCCAGTGGGATCATCCTGTCCGTATATG 648
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTGGTGGCCTTTGCCATGGCTTCTGTGGGCCAGTGGGATCATCCTGTCCGTATATG 300

Qy   649 AAGATCACCTGGAAGCTG 666
              ||||||||||||||||||
Db   301 AAGATCACCTGGAAGCTG 318

RESULT 7
AAS78869
ID   AAS78869 standard; cDNA; 1771 BP.
XX
AC   AAS78869;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #14673.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX

```
Qy   537  CTGCATGGAGTACAGCAGCAGGAGTCAGTCCTCGGGGTCCCCTCATGGTCCTGGTGGC  596
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  1102  CTGCATGGAGTACAGCAGCAGCAGGAGTCAGTCAGTGAGTCCCGGGCTGCCCCTCATGGTCCTGGTGGC  1161

Qy   597  CTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCAC  656
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162  CTTTGCCATTGGCTTCTGTGGGCCAGTGGGGATCATCCTGTCCTGCTATATGAAGATCAC  1221

Qy   657  CTGGAAGCTGTGCAGCACAGCT  678
          ||||||||||||||||||||||
Db  1222  CTGGAAGCTGTGCAGCACAGCT  1243
```

RESULT 8
AAS42840
ID   AAS42840 standard; cDNA; 536 BP.
XX
AC   AAS42840;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human G Protein-Coupled Receptor (GPCR) cDNA #35.
XX
KW   Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW   attention deficit disorder; anxiety; depression; bipolar disorder; ss;
KW   neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW   metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW   type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW   cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW   viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
KW   antidepressant; anorectic; PCR primer; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200162797-A2.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-US005676.
XX
PR   23-FEB-2000; 2000US-0184247P.
PR   23-FEB-2000; 2000US-0184303P.
PR   23-FEB-2000; 2000US-0184304P.
PR   23-FEB-2000; 2000US-0184305P.
PR   23-FEB-2000; 2000US-0184397P.
PR   02-MAR-2000; 2000US-0186457P.
PR   03-MAR-2000; 2000US-0186610P.
PR   09-MAR-2000; 2000US-0188064P.
PR   13-MAR-2000; 2000US-0188880P.
PR   03-APR-2000; 2000US-0194344P.
PR   23-JUN-2000; 2000US-0213861P.
PR   11-JUL-2000; 2000US-0217369P.
PR   11-JUL-2000; 2000US-0217370P.
PR   14-JUL-2000; 2000US-0218337P.
PR   20-JUL-2000; 2000US-0218492P.
XX
PA   (PHAA) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G, Wood LS, Parodi LA, Lind P;
XX
DR   WPI; 2001-570628/64.
DR   P-PSDB; AAU25588.
XX
PT   New isolated nucleic acid encoding a new G-protein coupled receptor
PT   polypeptide for detecting receptor modulators that can treat mental
PT   disorders, such as schizophrenia, anxiety, depression, or obesity.
XX
PS   Claim 4; Page 82-83; 279pp; English.
XX
CC   Sequences AAS42806-AAS42926 represent cDNA molecules and PCR primers for
CC   cDNA molecules encoding human G-protein coupled receptor (GPCR)
CC   polypeptides. The protein and DNA sequences of the invention can be used
CC   to identify compounds which bind to GPCR polypeptides and in screening

CC   for compounds that modulate GPCR activity. By screening a human subject
CC   for the presence of mutations in GPCR DNA, a GPCR-related disorder or a
CC   genetic predisposition can be diagnosed. The sequences can also be used
CC   for treatment and prevention of mental disorders such as schizophrenia,
CC   attention deficit disorder, anxiety, depression, dementia and bipolar
CC   disorder, neurological disorders such as Huntington's disease,
CC   Parkinson's disease and Tourette's syndrome, metabolic disorders such as
CC   obesity, anorexia and type 2 diabetes, cardiovascular disorders such as
CC   thrombosis, myocardial infarction, cardiomyopathy and atherosclerosis,
CC   viral infections caused by HIV and cancers
XX
SQ   Sequence 536 BP; 93 A; 166 C; 182 G; 95 T; 0 U; 0 Other;

Query Match   15.9%; Score 176; DB 4; Length 536;
Best Local Similarity   99.6%; Pred. No. 3.4e-76;
Matches 226;   Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   347 GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCC 406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGCTGTGGTCTGTGCCC 128

Qy   407 ACTGGGGCCCGGCCCTCCGCACGGCTGGCCGCCGCCAGGCTGTCGCTGGTGGCCATCTGGA 466
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   129 ACTGGGGCCCGGCCCTCCGCACGGCTGGCCGCCGCCAGGCTGTCGCTGGTGGCCATCTGGA 188

Qy   467 CCTTGGTGCTGCGACAGCGATGCCCCTGCTCTTGATGCCCATGACCAAGCCGCTGGTGG 526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   189 CCTTGGTGCTGCGACAGCGATGCCCCTGCTCTTGATGCCCATGACCAAGCCGCTGGTGG 248

Qy   527 GCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG 573
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   249 GCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG 295
```

RESULT 9
AAS30791/c
ID   AAS30791 standard; cDNA; 556 BP.
XX
AC   AAS30791;
XX
DT   04-DEC-2001 (first entry)
XX
DE   Human cDNA encoding G protein-coupled receptor nGPCR-2343.
XX
KW   Human; G protein-coupled receptor; nGPCR-x; ss; antiviral; analgesic;
KW   cytostatic; cardiant; antidiabetic; anoretic; hypotensive; hypertensive;
KW   antiparkinsonian; nootropic; neuroprotective; antidepressant;
KW   viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
KW   cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
KW   obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW   atherosclerosis; Parkinson's disease; psychosis; neurological disorder;
KW   schizophrenia; migraine; major depression; anxiety; mental disorder;
KW   manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.
XX
OS   Homo sapiens.
XX
PN   WO200166750-A2.
XX
PD   13-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US007322.
XX
PR   08-MAR-2000; 2000US-0187581P.
PR   08-MAR-2000; 2000US-0187582P.
PR   08-MAR-2000; 2000US-0187714P.
PR   08-MAR-2000; 2000US-0187715P.
PR   08-MAR-2000; 2000US-0187825P.
PR   08-MAR-2000; 2000US-0187828P.
PR   08-MAR-2000; 2000US-0187829P.
PR   08-MAR-2000; 2000US-0187830P.
PR   08-MAR-2000; 2000US-0187833P.
PR   08-MAR-2000; 2000US-0187874P.
PR   08-MAR-2000; 2000US-0187928P.

PR   08-MAR-2000; 2000US-0187929P.
PR   08-MAR-2000; 2000US-0187930P.
PR   08-MAR-2000; 2000US-0180049P.
PR   08-MAR-2000; 2000US-0192294P.
XX
PA   (PHAA) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G, Wood LS;
XX
DR   WPI: 2001-535778/59.
DR   P-PSDB; AAU19222.
XX
PT   Isolated nucleic acid molecules encoding G protein-coupled receptors
PT   termed nGPCR-x, useful in the treatment and diagnosis of viral
PT   infections, cancers and mental disorders (e.g. Parkinson's disease and
PT   schizophrenia).
XX
PS   Claim 4; Page 204-205; 336pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules encoding G
CC   protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,
CC   polypeptides, and modulators may be used in the treatment of diseases and
CC   conditions such as infections, such as viral infections caused by HIV-1
CC   (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and
CC   cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC   anorexia, hypotension, hypertension, myocardial infarction,
CC   atherosclerosis), Parkinson's disease, and psychotic and neurological
CC   disorders, including schizophrenia, migraine, major depression, anxiety,
CC   mental disorder, manic depression, and dyskinesias, such as Huntington's
CC   disease or Tourette's Syndrome and many other diseases and syndromes
CC   listed in the specification. nGPCR-x polynucleotides and polypeptides, as
CC   well as nGPCR-x modulators, may also be used in diagnostic assays for
CC   such diseases or conditions. The present sequence encodes a G protein-
CC   coupled receptor of the invention
XX
SQ   Sequence 556 BP; 138 A; 139 C; 151 G; 128 T; 0 U; 0 Other;

Query Match   14.4%; Score 159; DB 4; Length 556;
Best Local Similarity   99.5%; Pred. No. 8.2e-68;
Matches 209; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   697 AGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTACGGCTGCTGATGCTGGTGGCCGTG 756
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 AGCAGGAAAGGACGGCCACTGGCGAGGCTGCCTGCTTACGGCTGCTGATGCTGGTGGCCGTG 497

Qy   757 GTGGTGCTGCTTCAGGCCCTCACCACCTCAACATCAAGCAGTTCATGGCCGAGAGGGATGCTC 816
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 GTGGTGCTGCTTCAGGCCCTCACCACCTCAACATCAAGCAGTTCATGGCCGAGAGGTATGCTC 437

Qy   817 CACCTGCCATCCTGTGCCGAGCCGGAGGGCTTCTTACTGTCCCTTCAGGCCCACCGTGGCC 876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 CACCTGCCATCCTGTGCCGAGCCGGAGGGCTTCTTTACTGTCCCTTCAGGCCCACCGTGGCC 377

Qy   877 CTCATGAACATGAACTGTGCCATTACCCCA 906
          ||||||||||||||||||||||||||||||
Db   376 CTCATGAACATGAACTGTGCCATTACCCCA 347

RESULT 10
AAS98061/c
ID   AAS98061 standard; DNA; 594 BP.
XX
AC   AAS98061;
XX
DT   12-MAR-2002 (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #18.
XX
KW   Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;

KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity; diabetes;
KW   hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US015332.
XX
PR   11-MAY-2000; 2000US-0203217P.
PR   18-MAY-2000; 2000US-0205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin receptor
PT   polypeptides useful for identifying modulators that are useful for
PT   treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,
PT   stroke.
XX
PS   Disclosure; Page 123; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a nucleic
CC   acid molecule that is at least 80% identical to the G protein-coupled
CC   (GPCR) polynucleotides included in the specification. Also included are
CC   probes based on the GPCR sequences (including antisense probes), a host
CC   cell comprising an expression vector comprising the GPCR sequence,
CC   antibodies raised against the polypeptides, and methods of identifying
CC   modulators of the polypeptides. The polypeptides are useful for
CC   identifying modulator compounds which function as modulators, activators,
CC   repressors, agonists or antagonists of the novel GPCR polypeptides
CC   including the GAL4 polypeptide. The antibodies and nucleic acid probes as
CC   described above can be used to detect the presence of the polypeptides
CC   and nucleic acids and are used to diagnose a variety of diseases or
CC   disorders in which GPCRs are involved e.g., Alzheimer's disease,
CC   amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell
CC   carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic
CC   obstructive pulmonary disease, Crohn's disease, depression, epilepsy,
CC   macular degeneration, lymphoma, melanoma, multiple sclerosis,
CC   osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid
CC   arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other
CC   diseases listed in the specification. The probes and antibodies are also
CC   useful for diagnosing cognition and memory disorders, anorexia, hormonal
CC   release disorders, cardiovascular activity disorders, pain perception
CC   disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds
CC   that decrease or increase the expression of galanin receptor (GAL4) can
CC   be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR
CC   nucleic acid is useful for treating the above mentioned disorders by gene
CC   therapy techniques. The present sequence is a novel GPCR polynucleotide
CC   of the invention
XX
SQ   Sequence 594 BP; 164 A; 149 C; 171 G; 110 T; 0 U; 0 Other;

Query Match   9.3%; Score 103; DB 6; Length 594;
Best Local Similarity 100.0%; Pred. No. 3.4e-40;
Matches 103; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   290 CTTTCGGCAAGGGGCTCTGCAGGCTGCAGGCGGCGTTGCTCTTCACCGACACCTACGGGGG 349
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   530 CTTTCGGCAAGGGGCTCTGCAGGCTGCAGGCGGCGTTGCTCTTCACCGACACCTACGGGGG 471

Qy   350 GGGTCTCACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGC 392
          ||||||||||||||||||||||||||||||||||||||||||||
Db   470 GGGTCTCACCTCATGGCCTGTGTGAGCGTGGACCATTACCCAGC 428

```
RESULT 11
ABA15489
ID   ABA15489 standard; DNA; 22760 BP.
XX
AC   ABA15489;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 7820.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001334.
XX
PR   31-JAN-2000; 2000US-0179065P.         PR   08-SEP-2000; 2000US-0231243P.
PR   04-FEB-2000; 2000US-0180628P.         PR   08-SEP-2000; 2000US-0231244P.
PR   24-FEB-2000; 2000US-0184664P.         PR   08-SEP-2000; 2000US-0231413P.
PR   02-MAR-2000; 2000US-0186350P.         PR   08-SEP-2000; 2000US-0231414P.
PR   16-MAR-2000; 2000US-0189874P.         PR   08-SEP-2000; 2000US-0232081P.
PR   17-MAR-2000; 2000US-0190076P.         PR   12-SEP-2000; 2000US-0231968P.
PR   18-APR-2000; 2000US-0198123P.         PR   14-SEP-2000; 2000US-0232397P.
PR   19-MAY-2000; 2000US-0205515P.         PR   14-SEP-2000; 2000US-0232398P.
PR   07-JUN-2000; 2000US-0209467P.         PR   14-SEP-2000; 2000US-0232399P.
PR   28-JUN-2000; 2000US-0214486P.         PR   14-SEP-2000; 2000US-0232400P.
PR   30-JUN-2000; 2000US-0215135P.         PR   14-SEP-2000; 2000US-0232401P.
PR   07-JUL-2000; 2000US-0216647P.         PR   14-SEP-2000; 2000US-0233063P.
PR   07-JUL-2000; 2000US-0216880P.         PR   14-SEP-2000; 2000US-0233064P.
PR   11-JUL-2000; 2000US-0217487P.         PR   14-SEP-2000; 2000US-0233065P.
PR   11-JUL-2000; 2000US-0217496P.         PR   21-SEP-2000; 2000US-0234223P.
PR   14-JUL-2000; 2000US-0218290P.         PR   21-SEP-2000; 2000US-0234274P.
PR   26-JUL-2000; 2000US-0220963P.         PR   25-SEP-2000; 2000US-0234997P.
PR   26-JUL-2000; 2000US-0220964P.         PR   25-SEP-2000; 2000US-0234998P.
PR   14-AUG-2000; 2000US-0224518P.         PR   26-SEP-2000; 2000US-0235484P.
PR   14-AUG-2000; 2000US-0224519P.         PR   27-SEP-2000; 2000US-0235834P.
PR   14-AUG-2000; 2000US-0225213P.         PR   27-SEP-2000; 2000US-0235836P.
PR   14-AUG-2000; 2000US-0225214P.         PR   29-SEP-2000; 2000US-0236327P.
PR   14-AUG-2000; 2000US-0225266P.         PR   29-SEP-2000; 2000US-0236367P.
PR   14-AUG-2000; 2000US-0225267P.         PR   29-SEP-2000; 2000US-0236368P.
PR   14-AUG-2000; 2000US-0225268P.         PR   29-SEP-2000; 2000US-0236369P.
PR   14-AUG-2000; 2000US-0225270P.         PR   29-SEP-2000; 2000US-0236370P.
PR   14-AUG-2000; 2000US-0225447P.         PR   02-OCT-2000; 2000US-0236802P.
PR   14-AUG-2000; 2000US-0225757P.         PR   02-OCT-2000; 2000US-0237037P.
PR   14-AUG-2000; 2000US-0225758P.         PR   02-OCT-2000; 2000US-0237038P.
PR   14-AUG-2000; 2000US-0225759P.         PR   02-OCT-2000; 2000US-0237039P.
PR   18-AUG-2000; 2000US-0226279P.         PR   02-OCT-2000; 2000US-0237040P.
PR   22-AUG-2000; 2000US-0226681P.         PR   13-OCT-2000; 2000US-0239935P.
PR   22-AUG-2000; 2000US-0226868P.         PR   13-OCT-2000; 2000US-0239937P.
PR   22-AUG-2000; 2000US-0227182P.         PR   20-OCT-2000; 2000US-0240960P.
PR   23-AUG-2000; 2000US-0227009P.         PR   20-OCT-2000; 2000US-0241785P.
PR   30-AUG-2000; 2000US-0228924P.         PR   20-OCT-2000; 2000US-0241786P.
PR   01-SEP-2000; 2000US-0229287P.         PR   20-OCT-2000; 2000US-0241787P.
PR   01-SEP-2000; 2000US-0229343P.         PR   20-OCT-2000; 2000US-0241808P.
PR   01-SEP-2000; 2000US-0229344P.         PR   20-OCT-2000; 2000US-0241809P.
PR   01-SEP-2000; 2000US-0229345P.         PR   20-OCT-2000; 2000US-0241826P.
PR   05-SEP-2000; 2000US-0229509P.         PR   20-OCT-2000; 2000US-0242221P.
PR   05-SEP-2000; 2000US-0229513P.         PR   01-NOV-2000; 2000US-0244617P.
PR   06-SEP-2000; 2000US-0230437P.         PR   08-NOV-2000; 2000US-0246474P.
PR   06-SEP-2000; 2000US-0230438P.         PR   08-NOV-2000; 2000US-0246475P.
PR   08-SEP-2000; 2000US-0231242P.         PR   08-NOV-2000; 2000US-0246476P.
                                           PR   08-NOV-2000; 2000US-0246477P.
                                           PR   08-NOV-2000; 2000US-0246478P.
                                           PR   08-NOV-2000; 2000US-0246523P.
                                           PR   08-NOV-2000; 2000US-0246524P.
                                           PR   08-NOV-2000; 2000US-0246525P.
                                           PR   08-NOV-2000; 2000US-0246526P.
                                           PR   08-NOV-2000; 2000US-0246527P.
                                           PR   08-NOV-2000; 2000US-0246528P.
                                           PR   08-NOV-2000; 2000US-0246523P.
                                           PR   08-NOV-2000; 2000US-0246609P.
                                           PR   08-NOV-2000; 2000US-0246610P.
                                           PR   08-NOV-2000; 2000US-0246611P.
                                           PR   08-NOV-2000; 2000US-0246613P.
                                           PR   17-NOV-2000; 2000US-0249207P.
                                           PR   17-NOV-2000; 2000US-0249208P.
                                           PR   17-NOV-2000; 2000US-0249209P.
                                           PR   17-NOV-2000; 2000US-0249210P.
                                           PR   17-NOV-2000; 2000US-0249214P.
                                           PR   17-NOV-2000; 2000US-0249211P.
                                           PR   17-NOV-2000; 2000US-0249212P.
                                           PR   17-NOV-2000; 2000US-0249213P.
                                           PR   17-NOV-2000; 2000US-0249215P.
                                           PR   17-NOV-2000; 2000US-0249216P.
                                           PR   17-NOV-2000; 2000US-0249217P.
                                           PR   17-NOV-2000; 2000US-0249218P.
                                           PR   17-NOV-2000; 2000US-0249244P.
                                           PR   17-NOV-2000; 2000US-0249245P.
```

PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250391P.
PR   01-DEC-2000; 2000US-0251160P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system cancers
XX   and metastases.
XX
PS   Disclosure; SEQ ID NO 7820; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC   ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC   breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC   cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC   ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC   infectious diseases such as viral, bacterial, fungal and parasitic
CC   infections. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 22760 BP; 5353 A; 6538 C; 6154 G; 4715 T; 0 U; 0 Other;

Query Match   3.0%; Score 33; DB 5; Length 22760;
Best Local Similarity   100.0%; Pred. No. 1.2e-05;
Matches   33; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||||||
Db 12345 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 12377

RESULT 12
AAI22451
ID   AAI22451 standard; DNA; 232 BP.
XX
AC   AAI22451;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Probe #12384 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX

OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 12384; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;

Query Match   2.8%; Score 31; DB 4; Length 232;
Best Local Similarity   100.0%; Pred. No. 0.00011;
Matches   31; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   975 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||||
Db   91 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 121

RESULT 13
ABA67527
ID   ABA67527 standard; DNA; 232 BP.
XX
AC   ABA67527;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #15832.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.

```
PR    21-SEP-2000; 2000US-0234687P.
PR    27-SEP-2000; 2000US-0236359P.
PR    04-OCT-2000; 2000GB-0024263.
XX
PA    (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI    Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR    WPI; 2001-483447/52.
XX
PT    Human genome-derived single exon nucleic acid probes useful for analyzing
      gene expression in human fetal liver.
XX
PS    Claim 4; SEQ ID NO 15832; 639pp + Sequence Listing; English.
XX
CC    The invention relates to a single exon nucleic acid probe for measuring
CC    human gene expression in a sample derived from human foetal liver. The
CC    single exon nucleic acid probes may be used for predicting, measuring and
CC    displaying gene expression in samples derived from human fetal liver. The
CC    present sequence is a single exon nucleic acid probe of the invention.
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;

Query Match            2.8%;   Score 31;  DB 4;  Length 232;
Best Local Similarity  100.0%;  Pred. No. 0.00011;
Matches    31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    975 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||||
Db     91 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 121

RESULT 14
AAI47743
ID    AAI47743 standard; DNA; 232 BP.
XX
AC    AAI47743;
XX
DT    17-OCT-2001  (first entry)
XX
DE    Probe #16429 used to measure gene expression in human placenta sample.
XX
KW    Probe; microarray; human; placenta; antenatal diagnosis;
KW    genetic disorder; ss.
XX
OS    Homo sapiens.
XX
PN    WO200157272-A2.
XX
PD    09-AUG-2001.
XX
PF    30-JAN-2001; 2001WO-US000663.
XX
PR    04-FEB-2000; 2000US-0180312P.
PR    26-MAY-2000; 2000US-0207456P.
PR    30-JUN-2000; 2000US-00608408.
PR    03-AUG-2000; 2000US-00632366.
PR    21-SEP-2000; 2000US-0234687P.
PR    27-SEP-2000; 2000US-0236359P.
PR    04-OCT-2000; 2000GB-0024263.
XX
PA    (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI    Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR    WPI; 2001-488897/53.
XX
PT    Human genome-derived single exon nucleic acid probes useful for analyzing
PT    gene expression in human placenta.
XX
```

```
PS    Claim 25; SEQ ID NO 16429; 654pp; English.
XX
CC    The present invention relates to single exon nucleic acid probes (SENP).
CC    The present sequence is one such probe. The probes are useful for
CC    producing a microarray for predicting, measuring and displaying gene
CC    expression in samples derived from human placenta. The probes are useful
CC    for antenatal diagnosis of human genetic disorders
XX
SQ    Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;

Query Match            2.8%;   Score 31;  DB 4;  Length 232;
Best Local Similarity  100.0%;  Pred. No. 0.00011;
Matches    31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    975 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||||
Db     91 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 121

RESULT 15
ABA49623
ID    ABA49623 standard; DNA; 232 BP.
XX
AC    ABA49623;
XX
DT    01-FEB-2002  (first entry)
XX
DE    Human breast cell single exon nucleic acid probe #8318.
XX
KW    Human; microarray; single exon probe; gene expression; breast; disease;
KW    cancer; ss.
XX
OS    Homo sapiens.
XX
PN    WO200157271-A2.
XX
PD    09-AUG-2001.
XX
PF    30-JAN-2001; 2001WO-US000662.
XX
PR    04-FEB-2000; 2000US-0180312P.
PR    26-MAY-2000; 2000US-0207456P.
PR    30-JUN-2000; 2000US-00608408.
PR    03-AUG-2000; 2000US-0063366.
PR    21-SEP-2000; 2000US-0234687P.
PR    27-SEP-2000; 2000US-0236359P.
PR    04-OCT-2000; 2000GB-0024263.
XX
PA    (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI    Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR    WPI; 2001-496933/54.
XX
PT    New spatially-addressable set of single exon nucleic acid probes, useful
PT    for measuring gene expression in sample derived from human breast,
PT    comprises number of single exon nucleic acid probes.
XX
PS    Claim 4; SEQ ID NO 8318; 327pp + Sequence Listing; English.
XX
CC    The invention relates to a spatially-addressable set of single exon
CC    nucleic acid probes for measuring gene expression in a sample derived
CC    from human breast and BT 474 cells. The method involves contacting the
CC    probes with a collection of detectably labelled nucleic acids derived
CC    from mRNA of human breast, and then measuring the label bound to each
CC    probe of the microarray. The probes are useful for verifying the
CC    expression of regions of genomic DNA predicted to encode proteins. They
CC    are useful for gene discovery, and for determining predisposition and/or
CC    prognosing breast disease. Gene expression analysis is useful for
CC    assessing the toxicity of chemical agents on cells. The microarray of
CC    this invention presents a far greater diversity of probes for measuring
CC    gene expression, with far less bias than expressed sequence tag
CC    microarrays. The method is suitable for rapid production of functional
```

CC   information from genomic sequence. The present sequence is a single exon
CC   nucleic acid probe of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 232 BP; 12 A; 94 C; 5 G; 119 T; 0 U; 2 Other;

Query Match   2.8%; Score 31; DB 4; Length 232;
Best Local Similarity   100.0%; Pred. No. 0.00011;
Matches   31; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   975 TTCCTCCTCCTCCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||
Db   91 TTCCTCCTCCTCCCTCCTCCTCCTCCTCCACC 121

Search completed: February 21, 2004, 21:44:41
Job time : 487 secs

```
                    GenCore version 5.1.6

          Copyright (c) 1993 - 2004  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:     February 21, 2004, 21:28:30 ; Search time 109 Seconds
                              (without alignments)
                              5620.787 Million cell updates/sec

Title:      US-10-041-615-33
Perfect score: 1104
Sequence:   1 atgataaactaggcccta.........gtgtcacttgctttatgact 1104

Scoring table: OLIGO_NUC
            Gapop 60.0 , Gapext 60.0

Searched:   682709 seqs, 277475446 residues

Word size : 10

Total number of hits satisfying chosen parameters:     268057

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	31	2.8	6114	4	US-09-495-714C-5	Sequence 5, Appli
c 2	30	2.7	5822	3	US-08-899-595-4	Sequence 4, Appli
c 3	30	2.7	5822	3	US-08-899-595-5	Sequence 5, Appli
c 4	29	2.6	50	4	US-09-37-489-4	Sequence 4, Appli
c 5	29	2.6	1362	4	US-08-979-608A-12	Sequence 12, Appli
c 6	29	2.6	1362	4	US-09-517-849-12	Sequence 12, Appli
c 7	29	2.6	1362	4	US-09-616-289-12	Sequence 12, Appli
c 8	29	2.6	1422	4	US-08-979-608A-13	Sequence 13, Appli
c 9	29	2.6	1422	4	US-09-517-849-13	Sequence 13, Appli
c 10	29	2.6	1422	4	US-09-616-289-13	Sequence 13, Appli
c 11	29	2.6	1617	4	US-08-979-608A-11	Sequence 11, Appli
c 12	29	2.6	1617	4	US-09-517-849-11	Sequence 11, Appli
c 13	29	2.6	1617	4	US-09-616-289-11	Sequence 11, Appli
c 14	29	2.6	2561	4	US-09-616-289-48	Sequence 48, Appli
c 15	29	2.6	3728	1	US-08-111-939-1	Sequence 1, Appli
c 16	28	2.5	54	1	US-08-434-743-4	Sequence 4, Appli
c 17	28	2.5	58	3	US-08-860-038-16	Sequence 16, Appli
c 18	28	2.5	58	4	US-09-580-923-16	Sequence 16, Appli
c 19	27	2.4	30	1	US-08-068-747-4	Sequence 4, Appli
c 20	27	2.4	30	4	US-08-589-109A-12	Sequence 12, Appli
c 21	27	2.4	39	1	US-08-068-747-9	Sequence 9, Appli
c 22	27	2.4	48	4	US-09-580-923-34	Sequence 34, Appli
c 23	27	2.4	48	4	US-08-979-608A-36	Sequence 36, Appli
c 24	27	2.4	48	4	US-09-517-849-36	Sequence 36, Appli
c 25	27	2.4	48	4	US-09-616-289-36	Sequence 36, Appli
c 26	27	2.4	50	3	US-08-860-038-17	Sequence 17, Appli
c 27	27	2.4	50	4	US-09-580-923-17	Sequence 17, Appli

c 28	27	2.4	58	3	US-08-860-038-15	Sequence 15, Appl
c 29	27	2.4	58	4	US-09-580-923-15	Sequence 15, Appl
c 30	27	2.4	84	4	US-08-979-608A-37	Sequence 37, Appl
c 31	27	2.4	84	4	US-09-517-849-37	Sequence 37, Appl
c 32	27	2.4	84	4	US-09-616-289-37	Sequence 37, Appl
c 33	27	2.4	2421	4	US-09-187-330-4	Sequence 4, Appl
c 34	27	2.4	2487	4	US-09-187-330-54	Sequence 54, Appl
c 35	27	2.4	2556	4	US-09-817-310-1	Sequence 1, Appli
c 36	27	2.4	2721	6	US-5215881-2	Patent No. 5215881-2
c 37	27	2.4	5531	5	US-09-620-312D-619	Sequence 619, App
c 38	27	2.4	7475	2	US-08-971-036-1	Sequence 1, Appli
c 39	27	2.4	7475	1	US-09-096-570-1	Sequence 1, Appli
c 40	27	2.4	7475	4	US-09-265-617B-1	Sequence 1, Appli
c 41	27	2.4	8106	3	US-09-135-241-1	Sequence 1, Appli
c 42	27	2.4	8438	1	US-07-945-283-1	Sequence 1, Appli
c 43	27	2.4	10961	4	US-09-807-166-1	Sequence 1, Appli
c 44	27	2.4	51259	3	US-08-781-891-209	Sequence 209, App
c 45	27	2.4	51259	4	US-09-618-166-209	Sequence 209, App

```
                          ALIGNMENTS

RESULT 1
US-09-495-714C-5/c
; Sequence 5, Application US/09495714C
; Patent No. 6670465
; GENERAL INFORMATION:
;  APPLICANT: University Technologies International Inc.
;  TITLE OF INVENTION: RETINAL CALCIUM CHANNEL (ALPHA) 1F-SUBUNIT GENE
;  FILE REFERENCE: 45499.4 (formerly 45074.6)
;  CURRENT APPLICATION NUMBER: US/09/495,714C
;  CURRENT FILING DATE:  2000-02-01
;  NUMBER OF SEQ ID NOS: 138
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;  LENGTH: 6114
;  TYPE: DNA
;  ORGANISM: Mus musculus
US-09-495-714C-5

Query Match          2.8%;  Score 31; DB 4; Length 6114;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches  31;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
             |||||||||||||||||||||||||||||||
Db      2507 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 2477


RESULT 2
US-08-899-595-4
; Sequence 4, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
,  APPLICANT: Narumiya, Shuh
,  APPLICANT: Takahashi, No. 6111072uaki
,  TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
,  TITLE OF INVENTION: ENCODING SAME
,  NUMBER OF SEQUENCES: 14
,  CORRESPONDENCE ADDRESS:
,    ADDRESSEE: Foley & Lardner
,    STREET: 3000 K Street, N.W., Suite 500
,    CITY: Washington
,    STATE: D.C.
,    COUNTRY: USA
,    ZIP: 20007-5109
,  COMPUTER READABLE FORM:
,    MEDIUM TYPE: Floppy disk
,    COMPUTER: IBM PC compatible
,    OPERATING SYSTEM: PC-DOS/MS-DOS
,    SOFTWARE: PatentIn Release #1.0, Version #1.30
,  CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/899,595
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-242701
FILING DATE: 26-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 9-90170
FILING DATE: 25-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stephen A. Bent
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 049441/0112
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5822 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: CDS
LOCATION: 28..3972
US-08-899-595-4

Query Match   2.7%; Score 30; DB 3; Length 5822;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   976 TCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||
Db   1983 TCCTCCTCCTCCTCCTCCTCCTCCTCCACC 2012

RESULT 3
US-08-899-595-5/c
; Sequence 5, Application US/08899595
; Patent No. 611072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:

NAME: Stephen A. Bent
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 049441/0112
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 5822 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Human
US-08-899-595-5

Query Match   2.7%; Score 30; DB 3; Length 5822;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   976 TCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||
Db   3840 TCCTCCTCCTCCTCCTCCTCCTCCTCCACC 3811

RESULT 4
US-09-371-489-4/c
; Sequence 4, Application US/09371489
; Patent No. 6355803
; GENERAL INFORMATION:
; APPLICANT: Anand Natrajan
; APPLICANT: Qingping Jiang
; APPLICANT: David Sharpe
; APPLICANT: Say-Jong Law
; TITLE OF INVENTION: NEAR INFRARED CHEMILUMINESCENT
; TITLE OF INVENTION: ACRIDINIUM COMPOUNDS AND USES THEREOF
; FILE REFERENCE: CCDLT-258XX
; CURRENT APPLICATION NUMBER: US/09/371,489
; CURRENT FILING DATE: 1999-08-10
; EARLIER APPLICATION NUMBER: 60/096,073
; EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (223)...(223)
; OTHER INFORMATION: VANCO B PMP-PROBE 496.20 (ON PMP) IN EXAMPLE 16
US-09-371-489-4

Query Match   2.6%; Score 29; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.000119;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   974 CTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
          |||||||||||||||||||||||||||||
Db   29 CTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1

RESULT 5
US-08-979-608A-12/c
; Sequence 12, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;   Lees, Robert S.
;   Law, Simon W.

```
                    Arjona, Anibal A.
    TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
                        BINDING PROTEINS AND THEIR USES  IN DIAGNOSING AND
                        TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Fish & Richardson P.C.
    STREET: 225 Franklin Street
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02110-2804
COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: DOS
    SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/979,608A
    FILING DATE: 26-No. 6355451-1997
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 60/048,547
    FILING DATE: 03-JUN-1997
    APPLICATION NUMBER: US 60/031,930
    FILING DATE: 27-NOV-1996
ATTORNEY/AGENT INFORMATION:
    NAME: Myers, Louis
    REGISTRATION NUMBER: 35,965
    REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617/542-5070
    TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
    LENGTH: 1362 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
FEATURE:
    NAME/KEY:  Coding Sequence
    LOCATION:  1...696
    SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-08-979-608A-12

Query Match          2.6%;   Score 29;   DB 4;   Length 1362;
Best Local Similarity  100.0%;   Pred. No. 0.0002;
Matches    29;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
          |||||||||||||||||||||||||||||
Db     84 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 56


RESULT 6
US-09-517-849-12/c
; Sequence 12, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;            Lees, Robert S.
;            Law, Simon W.
;            Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
                      BINDING PROTEINS AND THEIR USES  IN DIAGNOSING AND
                      TREATING ATHEROSCLEROSIS
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Fish & Richardson P.C.
    STREET: 225 Franklin Street
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02110-2804
```

```
COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: DOS
    SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/517,849
    FILING DATE: 02-Mar-2000
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/979,608
    FILING DATE: 26-NOV-1997
ATTORNEY/AGENT INFORMATION:
    NAME: Myers, Louis
    REGISTRATION NUMBER: 35,965
    REFERENCE/DOCKET NUMBER: 10797-003001
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617/542-5070
    TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
    LENGTH: 1362 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
FEATURE:
    NAME/KEY:  Coding Sequence
    LOCATION:  1...696
    SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-517-849-12

Query Match          2.6%;   Score 29;   DB 4;   Length 1362;
Best Local Similarity  100.0%;   Pred. No. 0.0002;
Matches    29;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
          |||||||||||||||||||||||||||||
Db     84 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 56


RESULT 7
US-09-616-289-12/c
; Sequence 12, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(696)
US-09-616-289-12
```

Query Match   2.6%;   Score 29; DB 4; Length 1362;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCT 1000
           ||||||||||||||||||||||||||
Db   84 GTCTTCCTCCTCCTCCTCCTCCTCCT 56

RESULT 8
US-08-979-608A-13/c
; Sequence 13, Application US/0897960BA
; Patent No. 6355451
; GENERAL INFORMATION:
;   APPLICANT: Lees, Ann M.
;   Lees, Robert S.
;   Law, Simon W.
;   Arjona, Anibal A.
;   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;   TREATING ATHEROSCLEROSIS
;   NUMBER OF SEQUENCES: 42
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/979,608A
;   FILING DATE: 26-No- 6355451-1997
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/048,547
;   FILING DATE: 03-JUN-1997
;   APPLICATION NUMBER: US 60/031,930
;   FILING DATE: 27-NOV-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Myers, Louis
;   REGISTRATION NUMBER: 35,965
;   REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1422 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 1...756
;   SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-979-608A-13

Query Match   2.6%;   Score 29; DB 4; Length 1422;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCT 1000
           ||||||||||||||||||||||||||
Db   144 GTCTTCCTCCTCCTCCTCCTCCTCCT 116

RESULT 9
US-09-517-849-13/c
; Sequence 13, Application US/09517849

Patent No. 6605588
; GENERAL INFORMATION:
;   APPLICANT: Lees, Ann M.
;   Lees, Robert S.
;   Law, Simon W.
;   Arjona, Anibal A.
;   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;   TREATING ATHEROSCLEROSIS
;   NUMBER OF SEQUENCES: 42
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/517,849
;   FILING DATE: 02-Mar-2000
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/979,608
;   FILING DATE: 26-NOV-1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Myers, Louis
;   REGISTRATION NUMBER: 35,965
;   REFERENCE/DOCKET NUMBER: 10797-003001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1422 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: Coding Sequence
;   LOCATION: 1...756
;   SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-517-849-13

Query Match   2.6%;   Score 29; DB 4; Length 1422;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches   29; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCT 1000
           ||||||||||||||||||||||||||
Db   144 GTCTTCCTCCTCCTCCTCCTCCTCCT 116

RESULT 10
US-09-616-289-13/c
; Sequence 13, Application US/0961 6289
; Patent No. 6632923
; GENERAL INFORMATION:
;   APPLICANT: Lees, Ann M.
;   APPLICANT: Lees, Robert S.
;   APPLICANT: Law, Simon W.
;   APPLICANT: Arjona, Anibal A.
;   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;   TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;   TITLE OF INVENTION: ATHEROSCLEROSIS
;   FILE REFERENCE: 10797-004001
;   CURRENT APPLICATION NUMBER: US/09/616,289
;   CURRENT FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 09/517,849
;   PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR APPLICATION FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR APPLICATION FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR APPLICATION FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(756)
US-09-616-289-13

Query Match   2.6%;   Score 29;   DB 4;   Length 1422;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
Matches   29;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
          |||||||||||||||||||||||||||||
Db   144 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 116

RESULT 11
US-08-979-608A-11/c
; Sequence 11, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;   Lees, Robert S.
;   Law, Simon W.
;   Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/979,608A
;   FILING DATE: 26-No. 6355451-1997
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/048,547
;   FILING DATE: 03-JUN-1997
;   APPLICATION NUMBER: US 60/031,930
;   FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Myers, Louis
;   REGISTRATION NUMBER: 35,965
;   REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1617 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-08-979-608A-11

Query Match   2.6%;   Score 29;   DB 4;   Length 1617;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
Matches   29;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
          |||||||||||||||||||||||||||||
Db   339 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 311

RESULT 12
US-09-517-849-11/c
; Sequence 11, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;   Lees, Robert S.
;   Law, Simon W.
;   Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/517,849
;   FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/979,608
;   FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Myers, Louis
;   REGISTRATION NUMBER: 35,965
;   REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1617 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-517-849-11

Query Match   2.6%;   Score 29;   DB 4;   Length 1617;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
Matches   29;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
          |||||||||||||||||||||||||||||
Db   339 GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 311

RESULT 13
US-09-616-289-11/c
; Sequence 11, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.

```
;   APPLICANT: Lees, Robert S.
;   APPLICANT: Law, Simon W.
;   APPLICANT: Arjona, Anibal A.
;   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;   TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;   TITLE OF INVENTION: ATHEROSCLEROSIS
;   FILE REFERENCE: 10797-004001
;   CURRENT APPLICATION NUMBER: US/09/616,289
;   CURRENT FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 09/517,849
;   PRIOR FILING DATE: 2000-03-02
;   PRIOR APPLICATION NUMBER: US 08/979,608
;   PRIOR FILING DATE: 1997-11-26
;   PRIOR APPLICATION NUMBER: US 60/031,930
;   PRIOR FILING DATE: 1996-11-27
;   PRIOR APPLICATION NUMBER: US 60/048,547
;   PRIOR FILING DATE: 1997-06-03
;   NUMBER OF SEQ ID NOS: 53
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 11
;   LENGTH: 1617
;   TYPE: DNA
;   ORGANISM: Oryctolagus cuniculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(951)
; US-09-616-289-11

Query Match        2.6%;   Score 29;   DB 4;   Length 1617;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
Matches    29;   Conservative     0;   Mismatches     0;   Indels     0;   Gaps     0;

Qy      972  GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
             |||||||||||||||||||||||||||||
Db      339  GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 311

RESULT 14
US-09-616-289-48/c
; Sequence 48, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
;   APPLICANT: Lees, Ann M.
;   APPLICANT: Lees, Robert S.
;   APPLICANT: Law, Simon W.
;   APPLICANT: Arjona, Anibal A.
;   TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;   TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;   TITLE OF INVENTION: ATHEROSCLEROSIS
;   FILE REFERENCE: 10797-004001
;   CURRENT APPLICATION NUMBER: US/09/616,289
;   CURRENT FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 09/517,849
;   PRIOR FILING DATE: 2000-03-02
;   PRIOR APPLICATION NUMBER: US 08/979,608
;   PRIOR FILING DATE: 1997-11-26
;   PRIOR APPLICATION NUMBER: US 60/031,930
;   PRIOR FILING DATE: 1996-11-27
;   PRIOR APPLICATION NUMBER: US 60/048,547
;   PRIOR FILING DATE: 1997-06-03
;   NUMBER OF SEQ ID NOS: 53
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 48
;   LENGTH: 2561
;   TYPE: DNA
;   ORGANISM: Oryctolagus cuniculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (246)...(1895)
; US-09-616-289-48

Query Match        2.6%;   Score 29;   DB 4;   Length 2561;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
```

```
Matches    29;   Conservative     0;   Mismatches     0;   Indels     0;   Gaps     0;

Qy      972  GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1000
             |||||||||||||||||||||||||||||
Db     1283  GTCTTCCTCCTCCTCCTCCTCCTCCTCCT 1255

RESULT 15
US-08-111-939-1/c
; Sequence 1, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
;   APPLICANT: Kawai, Shinji
;   APPLICANT: Takeshita, Sunao
;   APPLICANT: Okazaki, Makoto
;   APPLICANT: Amann, Egon
;   TITLE OF INVENTION: Bone-Related Carboxypeptidase-like
;   TITLE OF INVENTION: Protein and Process for its Production
;   NUMBER OF SEQUENCES: 27
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;   ADDRESSEE: Dunner
;   STREET: 1300 I Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3315
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/111,939
;   FILING DATE: 26-AUG-1993
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 324033/92
;   FILING DATE: 03-DEC-1992
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 230029/92
;   FILING DATE: 28-AUG-1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Forman, David S.
;   REGISTRATION NUMBER: 33,694
;   REFERENCE/DOCKET NUMBER: 02481.1321-00000
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-408-4000
;   TELEFAX: 202-408-4000
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3728 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA to mRNA
;   ORIGINAL SOURCE:
;   ORGANISM: Mus musculus
;   STRAIN: osteoblastic cell line MC3T3E1
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 69..3452
; US-08-111-939-1

Query Match        2.6%;   Score 29;   DB 1;   Length 3728;
Best Local Similarity 100.0%;   Pred. No. 0.0002;
Matches    29;   Conservative     0;   Mismatches     0;   Indels     0;   Gaps     0;

Qy      973  TCTTCCTCCTCCTCCTCCTCCTCCTCCTC 1001
             |||||||||||||||||||||||||||||
Db     3385  TCTTCCTCCTCCTCCTCCTCCTCCTCCTC 3357
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: February 21, 2004, 21:36:35 ; Search time 468 Seconds
                                      (without alignments)
                                      8260.289 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaaactaggcccta.........gtgtcattgcttatgact 1104

Scoring table: OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched: 2308684 seqs, 1750822206 residues

Word size :   10

Total number of hits satisfying chosen parameters: 1853763

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   Published_Applications_NA.*
                1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2: /cgn2_6/ptodata/2/pubpna/US07_PCT_NEW_PUB.seq:*
                3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
               10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
               11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
               12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
               13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
               14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
               15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
               16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
               17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
               18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	1104	100.0	1104	15	US-10-041-615-33	Sequence 33, Appl
2	461	41.8	4319	14	US-10-017-161-1619	Sequence 1619, Ap
3	461	41.8	4319	15	US-10-292-798-1293	Sequence 1293, Ap
4	409	37.0	1526	9	US-09-728-952-40	Sequence 40, Appl
5	176	15.9	536	10	US-09-791-932-35	Sequence 35, Appl
6	159	14.4	556	11	US-09-801-944B-64	Sequence 64, Appl
7	103	9.3	594	9	US-09-812-102-90	Sequence 90, Appl
8	31	2.8	232	9	US-09-864-761-19928	Sequence 19928, A
9	31	2.8	384	9	US-09-864-761-3149	Sequence 3149, Ap
10	31	2.8	534	15	US-10-260-238-5532	Sequence 5532, Ap
11	31	2.8	2000	15	US-10-260-238-1941	Sequence 1941, Ap
12	31	2.8	2764	9	US-09-893-666A-1	Sequence 1, Appli
13	30	2.7	123	9	US-09-864-761-27614	Sequence 27614, A
14	30	2.7	180	14	US-10-029-386-17166	Sequence 17166, A
15	30	2.7	207	9	US-09-864-761-22442	Sequence 22442, A

16	30	2.7	276	9	US-09-864-761-17651	Sequence 17651, A
17	30	2.7	331	9	US-09-728-446-1021	Sequence 1021, Ap
c 18	30	2.7	393	10	US-09-803-719-2236	Sequence 2236, Ap
19	30	2.7	449	9	US-09-864-761-10977	Sequence 10977, A
20	30	2.7	464	9	US-09-864-761-871	Sequence 871, App
21	30	2.7	471	9	US-09-864-761-5679	Sequence 5679, Ap
c 22	30	2.7	504	14	US-10-029-386-6520	Sequence 6520, Ap
c 23	30	2.7	558	14	US-10-029-386-20250	Sequence 20250, A
24	30	2.7	598	14	US-10-029-386-3466	Sequence 3466, Ap
25	30	2.7	661	15	US-10-027-632-273400	Sequence 273400,
26	30	2.7	736	14	US-10-029-386-20317	Sequence 20317, A
c 27	30	2.7	739	12	US-10-424-599-113576	Sequence 113576,
c 28	30	2.7	2056	14	US-10-007-926A-264	Sequence 264, App
29	30	2.7	2068	15	US-10-291-172-24	Sequence 24, Appl
30	30	2.7	2079	15	US-10-291-172-23	Sequence 23, Appl
31	30	2.7	2095	13	US-10-044-090-665	Sequence 665, App
32	30	2.7	2098	15	US-10-291-172-25	Sequence 25, Appl
33	30	2.7	2120	14	US-10-119-428-26	Sequence 26, Appl
34	30	2.7	2129	14	US-10-119-428-25	Sequence 25, Appl
35	30	2.7	2150	14	US-10-119-428-27	Sequence 27, Appl
36	30	2.7	2211	12	US-10-424-599-44115	Sequence 4415, A
37	30	2.7	2317	15	US-10-291-172-399	Sequence 399, App
38	30	2.7	2317	15	US-10-291-172-400	Sequence 400, App
c 39	30	2.7	2317	15	US-10-291-172-401	Sequence 401, App
c 40	30	2.7	3953	14	US-10-232-561-11	Sequence 11, Appl
41	30	2.7	4450	9	US-09-764-887-497	Sequence 497, App
42	30	2.7	4450	14	US-10-073-961-497	Sequence 497, App
43	30	2.7	4454	9	US-09-764-887-496	Sequence 496, App
44	30	2.7	4454	14	US-10-073-961-496	Sequence 496, App
c 45	30	2.7	16033	14	US-10-311-455-1377	Sequence 1377, Ap

ALIGNMENTS

RESULT 1
US-10-041-615-33
; Sequence 33, Application US/10041615
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glenda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US2004001403BA1el GPCR-Like Proteins and Nucleic Acids Encod
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 33
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1080)
US-10-041-615-33

Query Match   100.0%; Score 1104; DB 15; Length 1104;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ATGATAAAACTAGGCCCTTACTTCACCACCCCACAAAAATAAAAACTAAGATTAAAGAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
RESULT 2
US-10-017-161-1619
; Sequence 1619, Application US/10017161
; Publication No. US20030143368A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT: ASAI, KIYOSHI
;  APPLICANT: AKIYAMA, YUTAKA
;  APPLICANT: ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1619
;  LENGTH: 4319
;  TYPE: DNA
;  ORGANISM: Homo sapiens
; FEATURE:
;  NAME/KEY: source
;  LOCATION: (1)..(4319)
; FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (201)..(308)
; FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (3128)..(3531)
; FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (3660)..(4119)
; FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (1)..(48)
;  OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1619

Query Match             41.8%;  Score 461;  DB 14;  Length 4319;
Best Local Similarity   99.8%;  Pred. No. 2.3e-225;
Matches  581;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;
```

```
Db            1 ATGATAAAACTAGGCCCTTACTTCACACCACCCCACAAAAATAAAAACTAAGATTAAAGAT 60

Qy           61 TTAAATGTGGAAAGCCAAACTGAAAATTTTTGGCGCTCTGTTCTCTCACAGCCCTC 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 TTAAATGTGGAAAGCCAAACTGTGAAAATTTTTGCGCTCTCGTTCTCTCACACAGCCCTC 120

Qy          121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGACAGG 180
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 CTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTTTGCCTTACCTGTCAAAAGAGACAGG 180

Qy          181 AAGATCAACTGCACAGGCATCTACCTGGTGACACCTGGCTGTGTCTGACCTGCTGTTCACC 240
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 AAGATCAACTGCACAGGCATCTACCTGGTGACACCTGGCTGTGTCTGACCTGCTGTTCACC 240

Qy          241 GTGGCCTTACCGGGAAGGGTGGTGTTATGTGTCGGGCTCCAGCTGGCCTTTCGGCAAG 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GTGGCCTTACCGGGAAGGGTGGTGTTATGTGTCGGGCTCCAGCTGGCCTTTCGGCAAG 300

Qy          301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACACACCTACGGGGGGGTCTACCTC 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCGACACACCTACGGGGGGGTCTACCTC 360

Qy          361 ATGGCCTGTGTGAGGTGTGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGCCGCGC 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 ATGGCCTGTGTGTGAGGTGTGACCATTACCCAGCTGTGGTCTGTGCCCACTGGGGCCGCGGC 420

Qy          421 CTCCGCACGGCTGGGCGCGGCCGCGCTGGGTCTCGGTGGCCATCTGGACCTTGGTGCTGCTG 480
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 CTCCGCACGGCTGGGCGGCCGCGGCCGCGCTGGGTCTGGTGGCCATCTGGACCTTGGTGCTGCTG 480

Qy          481 CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGCAAGCTGGCCTGC 540
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 CAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGGGGCAAGCTGGCCTGC 540

Qy          541 ATGGAGTACAGCAGCATGAGAGTCAGTCCTCGGGCTGCCCCTCATGGTCCTGGTGGCCTTT 600
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 ATGGAGTACAGCAGCATGAGAGTCAGTCCTCGGGGCTGCCCCTCATGGTCCTGGTGGCCTTT 600

Qy          601 GCCATTGGCTTCTCGTGTGGGGATCATCCTGCTATATGAAGATCACCTGG 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 GCCATTGGCTTCTCGTGTGGGGATCATCCTGCTATATGAAGATCACCTGG 660

Qy          661 AAGCTGTGCAGCACAGCTCGGGAGACCCAGTGACCAGCAGGAGGACGCCACTGGCCGA 720
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 AAGCTGTGCAGCACAGCTCGGGAGACCCAGTGACCAGCAGGAGGACGCCACTGGCCGA 720

Qy          721 GGCTGCTGCCTGCTTACCGGTCGCTGATGCGTGGTGGCCGTGGTGCTCGTGCTCTGTGCCACC 780
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GGCTGCTGCCTGCTTACCGGTCGCTGATGCGTGGTGGCCGTGGTGCTCGTGCTCTGTGCCACC 780

Qy          781 CTCAACATCAAGCAGTTCATGGCGAGAGGATGCTCCACCTGCTCTGCCGAGCGG 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 CTCAACATCAAGCAGTTCATGGCGAGAGGATGCTCCACCTGCTCTGCCGAGCGGG 840

Qy          841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCGCCACCTGCTCATGGAACATGTGGCATT 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 AGGGCTTTCTTACTGTCCCTTCAGGCCACCGCCACCTGCTCATGGAACATGTGGCATT 900

Qy          901 ACCCAATCATTTACTTCTTTGCATCCACCATTACAGGAAAATGCTCCTGGGCATTTTA 960
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 ACCCCAATCATTTACTTCTTTGCATCCACCATTACAGGAAAATGCTCCTGGGCATTTTA 960

Qy          961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCAGGAAAAGCTTCT 1020
                |||||||||||||||||||||||||||||||||||||||||||
Db          961 AAGCTCAAAGGGTCTTCCTCCTCCTCCTCCAGGAAAAGCTTCT 1020

Qy         1021 TCAGAAACACCAAGTATCACCCAGGCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC 1080
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 TCAGAAACACCAAGTATCACCCAGGCAGAGGTTCTATGTTCTTAGCGGAGCACGTGGTC 1080

Qy         1081 TAACGTGTCATTTGCTTTATGACT 1104
                ||||||||||||||||||||||||
Db         1081 TAACGTGTCATTTGCTTTATGACT 1104
```

```
Qy           97 CGGCTCCTCGTTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 156
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         2940 CGCTCTCGTTCTCTCACACAGCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT 2999

Qy          157 TGCCTTACCTGTCAAAAGAGACAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 216
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3000 TGCCTTACCTGTCAAAAGAGACAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG 3059

Qy          217 GCTGTGTCTGACCTGCTGTTCACCGGGAAGGGTGGTGTTATGTGCTG 276
                |||||||||||||||||||||||||||||||||||||||||||||||
Db         3060 GCTGTGTCTGACCTGCTGTTCACCGGGGAAGGGTGGTGTTATGTGCTG 3119

Qy          277 GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACC 336
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3120 GGCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCCTACACC 3179

Qy          337 GACACCTACGGGGGGGTCTACCTCATGGCCTGTGTGTGAGCGGTGAGCGTGGGCCGCCCGTG 396
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3180 GACACCTAC-GGGGGGTCTACCTCATGGCCTGTGTGTGAGCGGTGAGCGTGGGCCGCCCGTG 3238

Qy          397 GTCTGTGCCCACTGGGGCCGCGGCCGCGGCCGCGCTGGGTCTCGGTCTGCTG 456
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3239 GTCTGTGCCCACTGGGGCCGCGGCCGCGGCCGCGCTGGGTCTCGGTCTGCTG 3298

Qy          457 GCCATCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG 516
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         3299 GCCATCTGGACCTTGGTGCTGCTGCAGACGATGCCCTTGCTCTTGATGCCCATGACCAAG 3358
```

```
Qy    517  CCGCTGGTGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCAGCATGGAGTCAGTCCTCGGGCTG  576
           ||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db   3359  CCGCTGGTGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCAGCAGTCCTCGGGCTG  3418

Qy    577  CCCCTCATGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGGATCATCCTG  636
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3419  CCCCTCATGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGGATCATCCTG  3478

Qy    637  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  678
           ||||||||||||||||||||||||||||||||||||||||||
Db   3479  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  3520
```

RESULT 3
US-10-292-798-1293
; Sequence 1293, Application US/10292798
; Publication No. US2003025833A1
; GENERAL INFORMATION:
;   APPLICANT: SUWA, MAKIKO
;   APPLICANT: ASAI, KIYOSHI
;   APPLICANT: AKIYAMA, YUTAKA
;   APPLICANT: ABURATANI, HIROYUKI
;   TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;   FILE REFERENCE: 084335/166
;   CURRENT APPLICATION NUMBER: US/10/292,798
;   CURRENT FILING DATE: 2002-11-13
;   PRIOR APPLICATION NUMBER: 10/017,161
;   PRIOR FILING DATE: 2001-12-18
;   PRIOR APPLICATION NUMBER: JP 2001-246789
;   PRIOR FILING DATE: 2001-06-18
;   NUMBER OF SEQ ID NOS: 2070
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1293
;   LENGTH: 4319
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: (1)..(4319)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(308)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3128)..(3531)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3660)..(4119)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (1)..(48)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1293

Query Match   41.8%;   Score 461;   DB 15;   Length 4319;
Best Local Similarity   99.8%;   Pred. No. 2.3e-225;
Matches 581;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   1;

```
Qy     97  CGCTCCGTGTCTACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  156
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2940  CGCTCTCCGTGTCTACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  2999

Qy    157  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGGTGCACCTG  216
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3000  TGCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGTGCACCTG  3059

Qy    217  GTCTGTGTCGACCTGCTGTTCACCGTGGCCTTACCGGTGGAAGGTGGTGTTATGTGCTG  276
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3060  GCTGTGTCGACCTGCTGTTCACCGTGGCCTTACCGGTGGAAGGTGGTGTTATGTGCTG  3119

Qy    277  GGCTTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGCAGCGCGTTTGTGCTCTACACC  336
           ||||||||||||||||||||||
```

RESULT 4
US-09-728-952-40
; Sequence 40, Application US/09728952
; Patent No. US2002011302A1
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y., Tom
;   APPLICANT: Zhou, Ping
;   APPLICANT: Goodrich, Ryle
;   APPLICANT: Liu, Chenghua
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Wang, Dunrui
;   APPLICANT: Yamazaki, Vicki
;   APPLICANT: Ujwal, Manusha L.
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: No. US2002011302A1el Nucleic Acids and
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: 799
;   CURRENT APPLICATION NUMBER: US/09/728,952
;   CURRENT FILING DATE: 2000-11-30
;   NUMBER OF SEQ ID NOS: 101
;   SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 40
;   LENGTH: 1526
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (684)..(1016)
US-09-728-952-40

Query Match   37.0%;   Score 409;   DB 9;   Length 1526;
Best Local Similarity   99.7%;   Pred. No. 9.8e-199;
Matches 579;   Conservative   0;   Mismatches   1;   Indels   1;   Gaps   1;

```
Qy     98  GCTCTCCGTGTCTACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422  GCTCTCCGTGTCTACACAGCCCCTCCTGGTCTTCAGTGCCCTGGGAAACATCCTTGCCCTT  481

Qy    158  GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGG  217
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482  GCCTTACCTGTCAAAAGAGCAGGAAGATCAACTGCACAGGCATCTACCTGG  541

Qy    218  CTGTGTCGACCTGCTGTTCACCGTGGCCTTACCGGTGGAAGGTGGTGTTATGTGCTGG  277
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542  CTGTGTCGACCTGCTGTTCACCGTGGCCTTACCGGTGGAAGGTGGTGTTATGTGCTGG  601
```

```
Db   3120  GGCTCCAGCTGGCCTTTCGGCAAGGGGCTTCTGCAGGCTGCGACGCGGCGTTTGTGCTCTACACC  3179

Db    337  GACACCTACGGGGGGGTCTTACCTCATGGCCTGTGTGAGCGTGGACCCATTACCCAGCTGTG  396

Db   3180  GACACCTAC-GGGGGGTCTTACCTCATGGCCTGTGTGAGCGTGGACGACCCATTACCCAGCTGTG  3238

Db    397  GTCTGTGCCCACTGGGGCCCGGGCCTTCCCCACGGCTGGCCTGCCCCAGGCTGGTCTGCGTG  456

Db   3239  GTCTGTGCCCACTGGGGCCCGGGCCTTCCCCACGGCTGGCCTGCCCCAGGCTGGTCTGCGTG  3298

Db    457  GCCATCTGGACCTTGGTCTGCTGCTGCGACGATGCCCTTGCTCTTGATGCCCATGACCAAG  516

Db   3299  GCCATCTGGACCTTGGTCTGCTGCTGCGACGATGCCCTTGCTCTTGATGCCCATGACCAAG  3358

Db    517  CCGCTGGTGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCAGCATGGAGTCAGTCCTCGGGCTG  576

Db   3359  CCGCTGGTGTGGGCAAGCTGGCCTGCATGGAGTACAGCAGCAGCAGTCCTCGGGCTG  3418

Db    577  CCCCTCATGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGGATCATCCTG  636

Db   3419  CCCCTCATGGTCCTGTGGCCTTTGCCATTGGCTTCTGTGGGGATCATCCTG  3478

Db    637  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  678

Db   3479  TCCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  3520
```

```
Qy    278  GCTCCAGCTGGCCTTCGGCAAGGGGCTCTGCAGGCTGACGGGGTTTGTGCTCTACACCG  337
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602  GCTCCAGCTGGCCTTTCGGCAAGGGGCTCTGCAGGCTGACGGCGTTTGTGCTCTACACCG  661

Qy    338  ACACCTACGGGGGGTCTACCTCATGGCCTGTCGTGAGCGTGGACCATTACCCAGCTGTGG  397
                |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    662  ACACCTAC-GGGGGTCTACCTCATGGCCTGTCGTGAGCGTGGACCATTACCCAGCTGTGG  720

Qy    398  TCTGTGCCCACTGGGGCCCGCCGCCTCCGCACGGCTGGCCCAGGCTGGTCTGCGTGG  457
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  TCTGTGCCCACTGGGGCCCGCCGCCTCCGCACGGCTGGCCCAGGCTGGTCTGCGTGG  780

Qy    458  CCATCTGGACCTTGGTGCTGCTCGCAGAGATGCCCTTGCTCTTGATGCCCATGACCAAGC  517
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  CCATCTGGACCTTGGTGCTGCTCGCAGAGATGCCCTTGCTCTTGATGCCCATGACCAAGC  840

Qy    518  CGCTGGTGGGCAAGCTGGCCTGACCTGCATGGCAGCTACAGCAGCATGAGTCAGTCTGGGGCTGC  577
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CGCTGGTGGGCAAGCTGGCCTGACCTGCATGGCAGCTACAGCAGCATGAGTCAGTCTGGGGCTGC  900

Qy    578  CCCTCATGGTTCCTGGTGGCTTTGTGGGCCAGTGGGATCATCCTGT  637
                |||||||||||||||||||||||||||||||||||||||||||||
Db    901  CCCTCATGGTTCCTGGTGGCTTTGTGGGCCAGTGGGATCATCCTGT  960

Qy    638  CCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  678
                |||||||||||||||||||||||||||||||||||||||||
Db    961  CCTGCTATATGAAGATCACCTGGAAGCTGTGCAGCACAGCT  1001
```

RESULT 5
US-09-791-932-35

```
Query Match             15.9%;  Score 176;  DB 10;  Length 536;
Best Local Similarity   99.6%;  Pred. No. 1.8e-79;
Matches  226;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    347  GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGAGCCATTACCCAGCTGTGGTCTGTGCCC  406
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69   GGGGGGTCTACCTCATGGCCTGTGTGAGCGTGGAGCCATTACCCAGCTGTGGTCTGTGCCC  128

Qy    407  ACTGGGGCCCGACCCGCCTTCCGCACGGCTGGCCCGCGCCAGGCTGGTCTGCGTGGCCATCTGA  466
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129  ACTGGGGCCCGACCCGCCTTCCGCACGGCTGGCCCGCGCCAGGCTGGTCTGCGTGGCCATCTGA  188

Qy    467  CCTTGGTGCTGCTGCGAGACGATGCCCTTGCTTCTTGATGCCCATGACCAAGCCGCTGGTGG  526
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189  CCTTGGTGCTGCTGCGAGACGATGCCCTTGCTCTTGATGCCCATGACCAAGCCGCTGGTGG  248

Qy    527  GCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG  573
                |||||||||||||||||||||||||||||||||||||||||||||||
Db    249  GCAAGCTGGCCTGCATGGAGTACAGCAGCATGGAGTCAGTCCTCGGG  295
```

RESULT 6
US-09-801-944B-64/c

Query Match   14.4%; Score 159; DB 11; Length 556;
Best Local Similarity 99.5%; Pred. No. 9e-71;
Matches 209; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   697 AGGAGGAAAGGACGCCACTGGGCGAGGCTGCCTTACGCTGCTGATGCTGGTGGCCGTG 756
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 AGCAGGAAAGGACGACGCCACTGGGCGAGGCTGCCTTACGCTGCTGATGCTGGTGGCCGTG 497

Qy   757 GTGGTCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGCGCGAGAGGGATGCTC 816
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   496 GTGGTCTGCTTCAGCCCCTACCACCTCAACATCAAGCAGTTCATGCGCGAGAGGTATGCTC 437

Qy   817 CACCTGCCATCCTGTGCCGAGGCGAGGGCTTTCTTACTGTCCCTTCAGGCCACCGTGGCC 876
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 CACCTGCCATCCTGTGCCGAGGCGAGGGCTTTTCTTACTGTCCCTTCAGGCCACCGTGGCC 377

Qy   877 CTCATGAACATGAACTGTGGCATTACCCCA 906
          |||||||||||||||||||||||||||||
Db   376 CTCATGAACATGAACTGTGGCATTACCCCA 347

RESULT 7
US-09-812-102-90/c
; Sequence 90, Application US/09812102
; Patent No. US2002005179A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E
; TITLE OF INVENTION: No. US2002005179A1e1 G-Protein Coupled Receptor Homologs
; FILE REFERENCE: 5800-41 035800/183478
; CURRENT APPLICATION NUMBER: US/09/812,102
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/364,769
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 90
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;   OTHER INFORMATION: GPCR-RHODOPSIN
US-09-812-102-90

Query Match   9.3%; Score 103; DB 9; Length 594;
Best Local Similarity 100.0%; Pred. No. 4.1e-42;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   290 CTTTCGGCAAGGGGCTCTGCGAGGCTGCACGGCGGTTGTGCTCTACACCGACACCTACGGGGG 349
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   530 CTTTTCGGCAAGGGGCTCTGCGAGGCTGACGGCGGTTGTGCTCTACACCGACACCTACGGGG 471

Qy   350 GGGTCTCACCTCATGGCCTGTGTGAGCGTGACCATTACCCAGC 392
          ||||||||||||||||||||||||||||||||||||||||||
Db   470 GGGTCTCACCTCATGGCCTGTGTGAGCGTGACCATTACCCAGC 428

RESULT 8
US-09-864-761-19928
; Sequence 19928, Application US/09864761
; Patent No. US2002004876A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SEQ ID NO 19928
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; LENGTH: 232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;   OTHER INFORMATION: MAP TO AP000082.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.7
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; NAME/KEY: unsure
; LOCATION: 124
; NAME/KEY: unsure
; LOCATION: 127
US-09-864-761-19928

Query Match   2.8%; Score 31; DB 9; Length 232;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   975 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          |||||||||||||||||||||||||||||||||
Db   91 TTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 121

RESULT 9
US-09-864-761-3149
; Sequence 3149, Application US/09864761
; Patent No. US2002004876A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

```
;   APPLICANT:  Hanzel, David K.
;   APPLICANT:  Chen, Wensheng
;   TITLE OF INVENTION:  HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBBS USEFUL FOR
;   TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE:  Aeomica-X-1
;   CURRENT APPLICATION NUMBER:  US/09/864,761
;   CURRENT FILING DATE:  2001-05-23
;   PRIOR APPLICATION NUMBER:  US 60/180,312
;   PRIOR FILING DATE:  2000-02-04
;   PRIOR APPLICATION NUMBER:  US 60/207,456
;   PRIOR FILING DATE:  2000-05-26
;   PRIOR APPLICATION NUMBER:  US 09/632,366
;   PRIOR FILING DATE:  2000-08-03
;   PRIOR APPLICATION NUMBER:  GB 24263.6
;   PRIOR FILING DATE:  2000-10-04
;   PRIOR APPLICATION NUMBER:  US 60/236,359
;   PRIOR FILING DATE:  2000-09-27
;   PRIOR APPLICATION NUMBER:  PCT/US01/00666
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00667
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00664
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00669
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00665
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00568
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00663
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00662
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00661
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  PCT/US01/00670
;   PRIOR FILING DATE:  2001-01-30
;   PRIOR APPLICATION NUMBER:  US 60/234,687
;   PRIOR FILING DATE:  2000-09-21
;   PRIOR APPLICATION NUMBER:  US 09/608,408
;   PRIOR FILING DATE:  2000-06-30
;   PRIOR APPLICATION NUMBER:  US 09/774,203
;   PRIOR FILING DATE:  2001-01-29
;   NUMBER OF SEQ ID NOS:  49117
;   SOFTWARE:  Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 3149
;   LENGTH:  384
;   TYPE:  DNA
;   ORGANISM:  Homo sapiens
;   FEATURE:
;   OTHER INFORMATION:  MAP TO AP000082.1
;   OTHER INFORMATION:  EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
;   OTHER INFORMATION:  EXPRESSED IN BONE MARROW, SIGNAL = 3.1
;   OTHER INFORMATION:  EXPRESSED IN LUNG, SIGNAL = 3.1
;   OTHER INFORMATION:  EXPRESSED IN BRAIN, SIGNAL = 2.8
;   OTHER INFORMATION:  EXPRESSED IN HELA, SIGNAL = 3.3
;   OTHER INFORMATION:  EXPRESSED IN HEART, SIGNAL = 3.7
;   OTHER INFORMATION:  EXPRESSED IN PLACENTA, SIGNAL = 3.7
;   OTHER INFORMATION:  EXPRESSED IN HBL100, SIGNAL = 3.3
;   OTHER INFORMATION:  EXPRESSED IN BT474, SIGNAL = 2.7
;   OTHER INFORMATION:  EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   NAME/KEY:  unsure
;   LOCATION:  360
;   NAME/KEY:  unsure
;   LOCATION:  363
US-09-864-761-3149

Query Match             2.8%;   Score 31;   DB 9;   Length 384;
Best Local Similarity   100.0%;   Pred. No. 2.9e-05;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    975 TTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          |||||||||||||||||||||||||||||||
Db    327 TTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 357

RESULT 10
US-10-260-238-5532/c
;   Sequence 5532, Application US/10260238
;   Publication No. US20040016025A1
;   GENERAL INFORMATION:
;   APPLICANT:  Budworth, Paul R.
;   APPLICANT:  Moughamer, Todd G.
;   APPLICANT:  Briggs, Steven P.
;   APPLICANT:  Cooper, Bret
;   APPLICANT:  Glazebrook, Jane
;   APPLICANT:  Goff, Stephen A.
;   APPLICANT:  Katagiri, Fumiyaki
;   APPLICANT:  Kreps, Joel
;   APPLICANT:  Provart, Nicholas
;   APPLICANT:  Ricke, Darrell
;   APPLICANT:  Zhu, Tong
;   TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;   FILE REFERENCE: 60111-NP
;   CURRENT APPLICATION NUMBER: US/10/260,238
;   CURRENT FILING DATE:  2002-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,448
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,277
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/370,620
;   PRIOR FILING DATE: 2002-04-04
;   NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 5532
;   LENGTH:  534
;   TYPE:  DNA
;   ORGANISM:  Zea mays
US-10-260-238-5532

Query Match             2.8%;   Score 31;   DB 15;   Length 534;
Best Local Similarity   100.0%;   Pred. No. 2.9e-05;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
          |||||||||||||||||||||||||||||||
Db    285 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 255

RESULT 11
US-10-260-238-1941/c
;   Sequence 1941, Application US/10260238
;   Publication No. US20040016025A1
;   GENERAL INFORMATION:
;   APPLICANT:  Budworth, Paul R.
;   APPLICANT:  Moughamer, Todd G.
;   APPLICANT:  Briggs, Steven P.
;   APPLICANT:  Cooper, Bret
;   APPLICANT:  Glazebrook, Jane
;   APPLICANT:  Goff, Stephen A.
;   APPLICANT:  Katagiri, Fumiyaki
;   APPLICANT:  Kreps, Joel
;   APPLICANT:  Provart, Nicholas
;   APPLICANT:  Ricke, Darrell
;   APPLICANT:  Zhu, Tong
;   TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
;   FILE REFERENCE: 60111-NP
;   CURRENT APPLICATION NUMBER: US/10/260,238
;   CURRENT FILING DATE:  2002-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,448
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/325,277
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 60/370,620
;   PRIOR FILING DATE: 2002-04-04
;   NUMBER OF SEQ ID NOS: 6077
;   SEQ ID NO 1941
```

```
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-1941

Query Match        2.8%;   Score 31;   DB 15;   Length 2000;
Best Local Similarity   100.0%;   Pred. No. 2.8e-05;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy      975  TTCTCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
             ||||||||||||||||||||||||||||||||
Db     1797  TTCCTCTCCTCCTCCTCCTCCTCCTCCTCCACC 1767


RESULT 12
US-09-893-666A-1/c
; Sequence 1, Application US/0989366A
; Patent No. US20020129386A1
; GENERAL INFORMATION:
; APPLICANT: YAMASHITA, ICHIRO
; TITLE OF INVENTION: High estrogen-sensitive medaka fish
; FILE REFERENCE: 21021TUS-620-7249-0
; CURRENT APPLICATION NUMBER: US/09/893,666A
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: JP 2000-247729
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 2764
; TYPE: DNA
; ORGANISM: Oryzias latipes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (211)..(1935)
; OTHER INFORMATION:
US-09-893-666A-1

Query Match        2.8%;   Score 31;   DB 9;   Length 2764;
Best Local Similarity   100.0%;   Pred. No. 2.7e-05;
Matches   31;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy      972  GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
             ||||||||||||||||||||||||||||||||
Db     1000  GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 970


RESULT 13
US-09-864-761-27614
; Sequence 27614, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27614
; LENGTH: 123
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005686.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.71
; OTHER INFORMATION: EST HUMAN HIT: AI863650.1, EVALUE 6.30e+00
; OTHER INFORMATION: NT HIT: AF059679.1, EVALUE 4.30e+00
US-09-864-761-27614

Query Match        2.7%;   Score 30;   DB 9;   Length 123;
Best Local Similarity   100.0%;   Pred. No. 0.0001;
Matches   30;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy      973  TCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
             ||||||||||||||||||||||||||||||
Db       10  TCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 39


RESULT 14
US-10-029-386-17166
; Sequence 17166, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17166
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR22 184 2.0
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
```

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
US-10-029-386-17166

Query Match   2.7%; Score 30; DB 14; Length 180;
Best Local Similarity   100.0%; Pred. No. 9.9e-05;
Matches   30; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 1002
           ||||||||||||||||||||||||||||||
Db   128 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCC 157

RESULT 15
US-09-864-761-22442
; Sequence 22442, Application US/09864761
; Patent No. US2002004876A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22442
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

; OTHER INFORMATION: MAP TO AC005809.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: NT HIT: AF020931.1, EVALUE 4.90e-01
; OTHER INFORMATION: EST_HUMAN HIT: AA703592.1, EVALUE 7.30e-01
US-09-864-761-22442

Query Match   2.7%; Score 30; DB 9; Length 207;
Best Local Similarity   100.0%; Pred. No. 9.8e-05;
Matches   30; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   976 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
           |||||||||||||||||||||||||||||||||
Db   99 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 128

Search completed: February 21, 2004, 23:55:09
Job time : 475 secs

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: February 21, 2004, 21:26:30 ; Search time 2795 seconds
                      (without alignments)
                      11795.294 Million cell updates/sec

Title: US-10-041-615-33
Perfect score: 1104
Sequence: 1 atgataaaactaggccctta.........gtgtcatttgctttatgact 1104

Scoring table: OLIGO_NUC
               Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size :   10

Total number of hits satisfying chosen parameters: 25639868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   EST:*
  1: em_estba:*
  2: em_esthum:*
  3: em_estin:*
  4: em_estmu:*
  5: em_estov:*
  6: em_estpl:*
  7: em_estro:*
  8: em_htc:*
  9: gb_est1:*
 10: gb_est2:*
 11: gb_htc:*
 12: gb_est3:*
 13: gb_est4:*
 14: gb_est5:*
 15: em_estfun:*
 16: em_estom:*
 17: em_gss_hum:*
 18: em_gss_inv:*
 19: em_gss_pln:*
 20: em_gss_vrt:*
 21: em_gss_fun:*
 22: em_gss_mam:*
 23: em_gss_mus:*
 24: em_gss_pro:*
 25: em_gss_rod:*
 26: em_gss_phg:*
 27: em_gss_vrl:*
 28: gb_gss1:*
 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	103	9.3	594	9	AI659965	AI659965 we64e01.x
	2	34	3.1	143	13	BY346000	BY346000
c	3	32	2.9	203	28	BH689531	BH689531 BOMHB51TR
c	4	32	2.9	240	10	AW532693	AW532693 UI-R-BS0-

No.		ID	Name				
c	5	BH542697	BOHLP40TR	271	28	2.9	32
	6	CD845470	RFO2.143F	506	14	2.9	32
	7	CD839500	RFO2.115J	507	14	2.9	32
	8	CD840534	RFO2.119L	539	14	2.9	32
c	9	BZ499662	BONU34TR	570	28	2.9	32
	10	BH667854	BOMMY77TF	613	28	2.9	32
c	11	BH661526	BOHVN92TF	628	28	2.9	32
c	12	BZ430821	BONAF76TF	646	28	2.9	32
	13	BH977513	odh46a04.	701	28	2.9	32
	14	BH938360	odf67h07.	702	28	2.9	32
c	15	CD874215	AZ03.101K	706	14	2.9	32
c	16	BH537878	BOGLR54TR	729	28	2.9	32
c	17	BZ453671	BONHZ87TR	769	28	2.9	32
c	18	BZ068671	1jr43h08.	786	28	2.9	32
	19	BH704785	BOHYQ12TF	812	28	2.9	32
	20	BZ504153	BONHE32TR	836	28	2.9	32
	21	AL299503	Tetraodon	989	28	2.9	32
	22	CK209771	FGAS02155	1121	14	2.9	32
c	23	CD942005	RBM_34 Ge	185	14	2.8	31
c	24	CG026240	PGAAA12TF	210	28	2.8	31
c	25	BH061458	RPCI-24-3	328	28	2.8	31
c	26	CF063130	QCU17g11.	364	14	2.8	31
	27	AZ039092	RPCI-23-3	515	28	2.8	31
	28	AZ039077	RPCI-23-3	552	28	2.8	31
c	29	CB334566	3529_1.37	570	14	2.8	31
c	30	CG441247	OG5EA46TV	638	29	2.8	31
	31	CB079907	hp78f08.b	661	14	2.8	31
	32	CA235347	SCCFL409	664	14	2.8	31
	33	CA195996	SCSBAD108	686	14	2.8	31
	34	AB055167	Mus muscu	767	29	2.8	31
c	35	CB668345	OSJNEd16C	785	14	2.8	31
	36	CNS04PSZ	Tetraodon	791	29	2.8	31
	37	CG047970	PUIDV14TB	819	29	2.8	31
	38	CG253509	PUSCH40TC	833	29	2.8	31
	39	CG132477	PUFYW39TB	880	29	2.8	31
c	40	CNS02Z1L	Tetraodon	915	29	2.8	31
	41	CNS03YO4	Tetraodon	927	29	2.8	31
	42	CNS02S29	Tetraodon	951	29	2.8	31
c	43	CNS02UFR	Tetraodon	953	29	2.8	31
	44	CK161859	FGAS01443	1177	14	2.8	31
c	45	AZ804535	2M0065I21	94	28	2.7	30

ALIGNMENTS

RESULT 1
AI659965/c
LOCUS   AI659965   594 bp   mRNA   linear   EST 18-DEC-1999
DEFINITION   we64e01.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone
            IMAGE:2345880 3' similar to SW:EBI2_HUMAN P32249 EBV-INDUCED G
            PROTEIN-COUPLED RECEPTOR 2 ;, mRNA sequence.
ACCESSION   AI659965
VERSION   AI659965.1   GI:4763535
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 594)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1014   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 408.
FEATURES   Location/Qualifiers
     source   1 ..594

```
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="IMAGE:2345880"
           /dev_stage="fetal"
           /lab_host="DH10B (phage-resistant)"
           /clone_lib="Soares_thymus_NHFTh"
           /note="Organ: thymus, pooled; Vector: pT7T3D-Pac
           (Pharmacia) with a modified polylinker; Site 1: Not I; Site 1: Not I -
           Site_2: Eco RI; 1st strand cDNA was primed wIth a Not I -
           oligo(dT) primer [5'
           TGTTACCAATCTGAAGTGGGACGCGGCCGCCAACGTTTTTTTTTTTTTTTTT 3'],
           double-stranded cDNA was ligated to Eco RI adaptors
           (Pharmacia), digested with Not I and cloned into the Not I
           and Eco RI sites of the modified pT7T3 vector. Library
           went through one round of normalization. Library
           constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match              9.3%;   Score 103;  DB 9;  Length 594;
Best Local Similarity  100.0%;   Pred. No. 2.3e-41;
Matches 103;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      290  CTTTCGCAAGGGGCCTCTGCAGAGGCTGAGGCGTTTGTGCTCTACACCGACACCTACGGGG  349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      530  CTTTCGCAAGGGGCCTCTGCGAGAGGCTGAGGCGTTTGTGCTCTACACCGACACCTACGGGG  471

Qy      350  GGGTCTACCTCATGGCCTGTGTGAGGCGTGGACCATTACCCAGC  392
             ||||||||||||||||||||||||||||||||||||||||||||
Db      470  GGGTCTACCTCATGGCCTGTGTGAGGCGTGGACCATTACCCAGC  428
```

RESULT 2
BY346000
LOCUS   BY346000   143 bp   mRNA   linear   EST 12-DEC-2002
DEFINITION BY346000 RIKEN full-length enriched, whole joints Mus musculus cDNA
            clone L730017L22 5', mRNA sequence.
ACCESSION   BY346000
VERSION   BY346000.1 GI:26575488
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 143)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Garibaldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tonita,M.,
            Verardo,R., Wagner,L., Wahlested,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE   Analysis of the mouse transcriptome based on functional annotation

---

            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11.(2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken.
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Vassilis Aidinis (Biomedical Sciences
            Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
            Fleming street 16672 Vari,Greece) whose assistance we gratefully
            acknowledge.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source   1..143
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="L730017L22"
                     /tissue_type="whole joints"
                     /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN

Query Match   3.1%; Score 34;   DB 13; Length 143;
Best Local Similarity 100.0%;   Pred. No. 7.2e-06;
Matches 34; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACCC 1006
             |||||||||||||||||||||||||||||||||||||
Db   98 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCACCC 131

RESULT 3
BH689531/c
LOCUS   BH689531   203 bp   DNA   linear   GSS 19-FEB-2002
DEFINITION BOMHB51TR BO_2_3 KB Brassica oleracea genomic clone BOMHB51,
            genomic survey sequence.
ACCESSION   BH689531
VERSION   BH689531.1 GI:18759968
KEYWORDS   GSS.
SOURCE   Brassica oleracea
  ORGANISM Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE   Whole genome shotgun sequencing of Brassica oleracea
JOURNAL   Unpublished (2001)
COMMENT   Other_GSSs: BOMHB51TF
          Contact: Chris Town
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA.
          Tel: 301-838-3523
          Fax: 301-838-0208
          Email: cdtown@tigr.org
          DNA is from a doubled haploid provided by Tom Osborn.
          Seq primer: TR
          Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..203
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone_lib="BOMHB51"
                     /clone="BO_2_3_KB"
                     /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match   2.9%; Score 32; DB 28; Length 203;
Best Local Similarity   100.0%; Pred. No. 8.3e-05;
Matches   32; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
          |||||||||||||||||||||||||||||||||
Db   137 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCA 106

RESULT 4
AW532693
LOCUS   AW532693   240 bp   mRNA   linear   EST 06-MAR-2000
DEFINITION UI-R-BS0-amg-a-06-0-UI.s1 UI-R-BS0 Rattus norvegicus cDNA clone
            UI-R-BS0-amg-a-06-0-UI 3', mRNA sequence.
ACCESSION   AW532693
VERSION   AW532693.1 GI:7175107
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 240)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized embryo at 13 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 4-123,
            >(GAA)n#Simple repeat 143-240, >L162_5end#LINE/L1
            Seq primer: M13 Forward

FEATURES   POLYA=Yes.
     source   Location/Qualifiers
                    1..240
                    /organism="Rattus norvegicus"
                    /mol_type="mRNA"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    /clone="UI-R-BS0-amg-a-06-0-UI"
                    /dev_stage="embryonic 13 dpc"
                    /lab_host="DH10B (Life Technologies)"
                    /clone_lib="UI-R-BS0"
                    /note="Vector: pT77D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS0
                    library is derived from 13 dpc whole embryo tissue. For a
                    detailed description of the library from which this clone
                    was derived, please visit our web site at
                    ratest.eng.uiowa.edu.
                    TAG_TISSUE=embryo at 13 dpc
                    TAG_LIB=UI-R-BS0
                    TAG_SEQ=AATCC"

ORIGIN

Query Match   2.9%; Score 32; DB 10; Length 240;
Best Local Similarity   100.0%; Pred. No. 8.5e-05;
Matches   32; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 1004
          |||||||||||||||||||||||||||||||||
Db   80 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 111

RESULT 5
BH542697/c
LOCUS   BH542697   271 bp   DNA   linear   GSS 14-DEC-2001
DEFINITION BOHLP40TR BOHL Brassica oleracea genomic clone BOHLP40, genomic
            survey sequence.
ACCESSION   BH542697
VERSION   BH542697.1 GI:17794478
KEYWORDS   GSS.
SOURCE   Brassica oleracea
ORGANISM   Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 271)
AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE   Whole genome shotgun sequencing of Brassica oleracea
JOURNAL   Unpublished (2001)
COMMENT   Other_GSSs: BOHLP40TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES   Location/Qualifiers
     source   1..271
                    /organism="Brassica oleracea"
                    /mol_type="genomic DNA"
                    /strain="TO1000DH3"
                    /db_xref="taxon:3712"
                    /clone="BOHLP40"
                    /clone_lib="BOHL"
                    /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
                    genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match   2.9%; Score 32; DB 28; Length 271;
Best Local Similarity   100.0%; Pred. No. 8.6e-05;
Matches   32; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
          ||||||||||||||||||||||||||||||||
Db    132 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 101

RESULT 6
CD845470
LOCUS       CD845470                 506 bp    mRNA    linear    EST 11-JUL-2003
DEFINITION  RFO2.143F14F011228 RFO2 Brassica napus cDNA clone RFO2143F14, mRNA
            sequence.
ACCESSION   CD845470
VERSION     CD845470.1  GI:32529292
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 506)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES    Location/Qualifiers
    source          1..506
                    /organism="Brassica napus"
                    /mol_type="mRNA"
                    /cultivar="samourai (restored line)"
                    /db_xref="taxon:3708"
                    /clone="RFO2143F14"
                    /tissue_type="anthers"
                    /clone_lib="RFO2"
ORIGIN

Query Match             2.9%;  Score 32;  DB 14;  Length 506;
Best Local Similarity   100.0%;  Pred. No. 9.6e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 1004
          ||||||||||||||||||||||||||||||||
Db     86 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 117

RESULT 7
CD839500
LOCUS       CD839500                 507 bp    mRNA    linear    EST 10-JUL-2003
DEFINITION  RFO2.115J02F010530 RFO2 Brassica napus cDNA clone RFO2115J02, mRNA
            sequence.
ACCESSION   CD839500
VERSION     CD839500.1  GI:32521440
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 507)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.fr).
            and http://genoplante-info.infobiogen.fr).
FEATURES    Location/Qualifiers
    source          1..507
                    /organism="Brassica napus"
                    /mol_type="mRNA"
                    /cultivar="samourai (restored line)"
                    /db_xref="taxon:3708"
                    /clone="RFO2115J02"
                    /tissue_type="anthers"
                    /clone_lib="RFO2"
ORIGIN

Query Match             2.9%;  Score 32;  DB 14;  Length 507;
Best Local Similarity   100.0%;  Pred. No. 9.6e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 1004
          ||||||||||||||||||||||||||||||||
Db     87 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 118

RESULT 8
CD840534
LOCUS       CD840534                 539 bp    mRNA    linear    EST 10-JUL-2003
DEFINITION  RFO2.119L18F010605 RFO2 Brassica napus cDNA clone RFO2119L18, mRNA
            sequence.
ACCESSION   CD840534
VERSION     CD840534.1  GI:32522474
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 539)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.fr).
            and http://genoplante-info.infobiogen.fr).
FEATURES    Location/Qualifiers
    source          1..539
                    /organism="Brassica napus"
                    /mol_type="mRNA"
                    /cultivar="samourai (restored line)"
                    /db_xref="taxon:3708"
                    /clone="RFO2119L18"
                    /tissue_type="anthers"
                    /clone_lib="RFO2"
ORIGIN

Query Match             2.9%;  Score 32;  DB 14;  Length 539;
Best Local Similarity   100.0%;  Pred. No. 9.7e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    973 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 1004
          ||||||||||||||||||||||||||||||||
Db    119 TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCAC 150

RESULT 9
BZ499662/c
LOCUS       BZ499662                 570 bp    DNA    linear    GSS 16-DEC-2002
DEFINITION  BONLU34TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONLU34,
            genomic survey sequence.
```

```
ACCESSION   BZ499662
VERSION     BZ499662.1  GI:27014967
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 570)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSs: BONLU34TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES    Location/Qualifiers
     source 1..570
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone_lib="BO_1.6_2_KB_tot"
            /clone="BONLU34"
            /note="Vector: pHOS1; Site 1: BstXI; 1.6-2 kb sheared
            total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 570;
Best Local Similarity  100.0%;  Pred. No. 9.7e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
          ||||||||||||||||||||||||||||||||
Db     63 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 32

RESULT 10
BH667854
LOCUS       BH667854             613 bp   DNA      linear   GSS 19-FEB-2002
DEFINITION  BOMMY77TF BO_2_3_KB Brassica oleracea genomic clone BOMMY77,
            genomic survey sequence.
ACCESSION   BH667854
VERSION     BH667854.1  GI:18727606
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 613)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
     source 1..613
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone_lib="BOMMY77"
            /clone_lib="BO_2_3_KB"
            /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 613;
Best Local Similarity  100.0%;  Pred. No. 9.9e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
          ||||||||||||||||||||||||||||||||
Db    509 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 540

RESULT 11
BH661526/c
LOCUS       BH661526             628 bp   DNA      linear   GSS 19-FEB-2002
DEFINITION  BOHVN92TF BO_2_3_KB Brassica oleracea genomic clone BOHVN92,
            genomic survey sequence.
ACCESSION   BH661526
VERSION     BH661526.1  GI:18720301
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 628)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
     source 1..628
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone_lib="BOHVN92"
            /clone_lib="BO_2_3_KB"
            /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 628;
Best Local Similarity  100.0%;  Pred. No. 9.9e-05;
Matches   32;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
          ||||||||||||||||||||||||||||||||
Db    509 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 540

RESULT 12
BZ430821/c
LOCUS       BZ430821             646 bp   DNA      linear   GSS 13-DEC-2002
DEFINITION  BONAF76TF BO_1_6_KB tot Brassica oleracea genomic clone BONAF76,
            genomic survey sequence.
ACCESSION   BZ430821
VERSION     BZ430821.1  GI:26676292
KEYWORDS    GSS.
SOURCE      Brassica oleracea
```

```
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 646)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSs: BONAF76TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES    Location/Qualifiers
     source  1..646
             /organism="Brassica oleracea"
             /mol_type="genomic DNA"
             /strain="TO1000DH3"
             /db_xref="taxon:3712"
             /clone="BONAF76"
             /clone_lib="BO_1.6_2_KB_tot"
             /note="Vector: pHOS; Site_1: BstXI; 1.6-2 kb sheared
             total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 646;
Best Local Similarity   100.0%;  Pred. No. 9.9e-05;  Indels 0;  Gaps 0;
Matches 32;  Conservative 0;  Mismatches 0;

Qy  972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
        ||||||||||||||||||||||||||||||
Db  138 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 107

RESULT 13
BH977513
LOCUS       BH977513   701 bp   DNA   linear   GSS 02-OCT-2002
DEFINITION  odh46a04.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH977513
VERSION     BH977513.1  GI:23460516
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 701)
AUTHORS     Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE       Whole genome shotgun reads from Brassica oleracea
JOURNAL     Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odh46   row: a   column: 04
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 65
            High quality sequence stop: 474.
FEATURES    Location/Qualifiers
     source  1..701
             /organism="Brassica oleracea"
             /mol_type="genomic DNA"
             /db_xref="taxon:3712"
             /clone_lib="B.oleracea002"
             /note="Vector: pOTw13; Whole genome shotgun library from
             flowering buds. DNA was purified  from a crude nuclear
             prep using Brassica oleracea TO1000DH3 buds  provided by
             Thosmas Osborn at the University of Wisconsin. Genomic
             DNA was provided by Pablo Rabinowicz (CSHL) and the
             shotgun library prepared at Washington University Genome
             Sequencing Center."
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 701;
Best Local Similarity   100.0%;  Pred. No. 0.0001;  Indels 0;  Gaps 0;
Matches 32;  Conservative 0;  Mismatches 0;

Qy  972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
        ||||||||||||||||||||||||||||||||
Db  143 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 174

RESULT 14
BH938360/c
LOCUS       BH938360   702 bp   DNA   linear   GSS 01-OCT-2002
DEFINITION  odf67h07.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH938360
VERSION     BH938360.1  GI:23418426
KEYWORDS    GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 702)
AUTHORS     Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
TITLE       Whole genome shotgun reads from Brassica oleracea
JOURNAL     Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: odf67   row: h   column: 07
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 18
            High quality sequence stop: 551.
FEATURES    Location/Qualifiers
     source  1..702
             /organism="Brassica oleracea"
             /mol_type="genomic DNA"
             /db_xref="taxon:3712"
             /clone_lib="B.oleracea002"
             /note="Vector: pOTw13; Whole genome shotgun library from
             flowering buds. DNA was purified  from a crude nuclear
             prep using Brassica oleracea TO1000DH3 buds provided by
             Thosmas Osborn at the University of Wisconsin. Genomic
             DNA was provided by Pablo Rabinowicz (CSHL) and the
             shotgun library prepared at Washington University Genome
             Sequencing Center."
ORIGIN

Query Match             2.9%;  Score 32;  DB 28;  Length 702;
Best Local Similarity   100.0%;  Pred. No. 0.0001;  Indels 0;  Gaps 0;
Matches 32;  Conservative 0;  Mismatches 0;

Qy  972 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 1003
        ||||||||||||||||||||||||||||||
Db  136 GTCTTCCTCCTCCTCCTCCTCCTCCTCCTCCA 105

RESULT 15
CD874215/c
LOCUS       CD874215   706 bp   mRNA   linear   EST 11-JUL-2003
DEFINITION  AZO3.101K13F010928 AZO3 Triticum aestivum cDNA clone AZO3101K13,
            mRNA sequence.
```

ACCESSION   CD874215
VERSION   CD874215.1 GI:3258031
KEYWORDS   EST.
SOURCE   Triticum aestivum (bread wheat)
ORGANISM   Triticum aestivum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 706)
AUTHORS   Genoplante.
TITLE   Genoplante, a major partnership french program in plant genomics
JOURNAL   Unpublished (2003)
COMMENT   Contact: Genoplante
             Genoplante
             93, rue Henri Rochefort 91025 EVRY CEDEX France
             Tel: 33 1 69 47 54 00
             Fax: 33 1 69 47 54 10
             This sequence has been generated in the framework of the french
             plant genomics programme 'Genoplante' (http://www.genoplante.com
             and http://genoplante-info.infobiogen.fr).
FEATURES   Location/Qualifiers
     source   1..706
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="recital"
                     /db_xref="taxon:4565"
                     /clone="AZ03101K13"
                     /tissue_type="leaf"
                     /clone_lib="AZ03"
ORIGIN

Query Match   2.9%; Score 32; DB 14; Length 706;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches   32; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   974 CTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC 1005
          ||||||||||||||||||||||||||||||||
Db   95 CTTCCTCCTCCTCCTCCTCCTCCTCCTCCACC   64

Search completed: February 21, 2004, 23:44:56
Job time : 2815 secs

OM protein - protein search, using sw model

Copyright GenCore version 5.1.6
(c) 1993 - 2004 Compugen Ltd.

Run on: February 19, 2004, 19:17:05 ; Search time 43 Seconds
(without alignments)
1328.874 Million cell updates/sec

Title: US-10-041-615-34
Perfect score: 1876
Sequence: 1 MIKLGPYFTPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq 19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	Length	DB ID	Description
1	1876	100.0	360	23		ABG32286	Human GPCR16 prote
2	1506	80.3	377	22		ABG14682	Novel human diagno
3	632.5	33.7	361	15		AAR54080	Epstein Barr virus
4	632.5	33.7	361	19		AAW53623	Epstein Barr virus
5	632.5	33.7	361	21		AAY90630	Human G protein-co
6	632.5	33.7	361	24		ABP81814	Human BBV-induced
7	630.5	33.6	361	21		AAY90664	Human mutant G pro
8	555	29.6	110	23		ABG66698	Human novel polype
9	417	22.2	177	22		AAU25588	Human G Protein-Co

10	395.5	21.1	339	21	AAY90652	Human mutant G pro
11	391.5	20.9	339	15	AAR53752	Seven transmembran
12	391.5	20.9	339	18	AAW07617	Human G-protein th
13	391.5	20.9	339	19	AAW48733	Human R12 seven tr
14	391.5	20.9	339	21	AAB21697	Human 7TM receptor
15	391.5	20.9	339	21	AAY90618	Human G protein-co
16	391.5	20.9	339	22	AAB82853	Human P2Y-like GPC
17	391.5	20.9	339	23	AAU98065	Human cysteinyl le
18	391.5	20.9	339	23	AAU91239	Human 7 transmembr
19	391.5	20.9	339	24	ABU11906	Human G-protein co
20	391.5	20.1	367	24	ABP81909	Human G-protein-co
21	377.5	20.1	177	22	AAU19222	Human G-protein-co
22	355	18.9	371	23	ABG35299	Human PAR1 type th
23	352.5	18.8	402	23	ABG35228	Human PAR1 type th
24	352.5	18.8	425	15	ABG35300	Human PAR1 type th
25	340	18.1	425	15	AAR60698	Fragment of the hu
26	339.5	18.1	425	13	AAR27240	Human thrombin rec
27	339.5	18.1	425	19	AAW51407	Human protease-act
28	339.5	18.1	425	20	AAY49570	Human thrombin rec
29	339.5	18.1	425	23	AAG80697	Human thrombin pro
30	339.5	18.1	425	23	AAE17032	Human thrombin pro
31	339.5	18.1	425	24	ABR47449	Breast cancer asso
32	339.5	18.1	425	24	ABP81919	Human thrombin rec
33	339.5	18.1	425	24	ABG73511	Human thrombin pro
34	339.5	18.1	426	21	AAY45035	Human thrombin rec
35	339.5	18.1	892	18	AAW16314	X. laevis thrombin
36	332.5	17.7	420	24	ABG73510	Human thrombin rec
37	332	17.7	537	23	AAU74538	Human P2Y purinoce
38	331	17.6	328	17	AAR91225	Human placenta G-p
39	329	17.5	328	22	AAE04393	Human P2-purinergi
40	328	17.5	328	24	ABP81869	Human purinergic r
41	328	17.5	355	22	ABB56343	Non-endogenous hum
42	324	17.3	328	18	AAW09433	Human placenta pur
43	324	17.3	355	18	AAW07618	Human G-protein re
44	324	17.3	355	19	AAW49807	Human G-protein ch
45	324	17.3	355	20	AAW97868	Human CC chemokine

ALIGNMENTS

RESULT 1
ABG32286
ID   ABG32286 standard; Protein; 360 AA.
XX
AC   ABG32286;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Human GPCR16 protein.
XX
KW   G protein coupled receptor related protein; human; GPCR;
KW   cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
KW   von Hippel-Lindau syndrome; Alzheimer's disease; tuberous sclerosis;
KW   hypercalcaemia; Parkinson's disease; Huntington's disease;
KW   cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW   ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW   pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW   inflammatory bowel disease; infertility; hypertension; scleroderma;
KW   arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW   infection; graft-versus-host disease.
XX
OS   Homo sapiens.
XX
PN   WO200264793-A2.
XX
PD   22-AUG-2002.
XX
PF   03-JAN-2002; 2002WO-US00056.
XX
PR   03-JAN-2001; 2001US-259552P.
PR   09-JAN-2001; 2001US-260544P.
PR   20-MAR-2001; 2001US-277405P.

```
Qy    301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
```

```
RESULT 2
ABG14682
ID   ABG14682 standard; Protein; 377 AA.
XX
AC   ABG14682;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #14673.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS78869.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 45041; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  377 AA;

Query Match             80.3%;  Score 1506;  DB 22;  Length 377;
Best Local Similarity   80.6%;  Pred. No. 6.8e-152;
Matches 295;  Conservative  1;  Mismatches  4;  Indels  66;  Gaps  1;

Qy     35 LFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGS 94
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Casman SJ, Edinger SR,  Ellerman K,  Smithson G,  Kekuda R;
PI   Muralidhara P;
XX
DR   WPI; 2002-643487/69.
DR   N-PSDB; ABS64723.
XX
PT   New isolated G protein coupled receptor polypeptides and
PT   polynucleotides, useful in gene therapy, particularly for treating or
PT   preventing cardiomyopathy, atherosclerosis, diabetes, multiple
PT   sclerosis, Crohn's disease or cancer -
XX
PS   Claim 1; Page 98; 252pp; English.
XX
CC   This invention relates to a new isolated G-protein coupled receptor
CC   (GPCRX) polypeptide sequence and the cDNA encoding it. The GPCRX
CC   polypeptide, GPCRX nucleic acid and an antibody specific to the protein
CC   are useful for treating, preventing or alleviating a GPCRX-associated
CC   disorder or a particular state in a subject, particularly a human. In
CC   particular, the disorder is cardiomyopathy, atherosclerosis, diabetes,
CC   or a disorder related to cell signal processing and metabolic pathway
CC   modulation. The GPCRX polypeptide and nucleic acid are also useful for
CC   diagnosing the presence of or predisposition to a disease associated
CC   with altered levels of GPCRX, particularly cancer. The GPCRX nucleic
CC   acid and polypeptide are especially useful in the manufacture of a
CC   medicament for therapeutic or prophylactic applications for disorders
CC   associated with aberrant GPCRX expression or activity, e.g. Von
CC   Hippel-Lindau syndrome, Alzheimer's disease, stroke, tuberous
CC   sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease,
CC   cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis,
CC   ataxia-telangiectasia, leukodystrophies, addiction, anxiety, depression,
CC   pain, obesity, Crohn's disease, osteoporosis, inflammatory bowel
CC   disease, infertility, hypertension, scleroderma, haemophilia,
CC   asthma, arthritis, human immunodeficiency virus; autoimmune disease;
CC   HIV, viral, fungal, bacterial or protozoal infections, or graft-versus
CC   -host disease. The DNA encoding the protein is useful in gene therapy
CC   for treating the above conditions. The polypeptides can be used as
CC   immunogens to produce antibodies and as vaccines. The nucleic acids are
CC   further used as hybridisation probes, in chromosome mapping, tissue
CC   typing, preventive medicine, and pharmacogenomics. These are also useful
CC   in developing powerful assay system for functional analysis of various
CC   human disorders, as well as in diagnostic applications. The present
CC   sequence represents a human G protein coupled receptor related protein
CC   (GPCR) of the invention.
XX
SQ   Sequence  360 AA;

Query Match             100.0%;  Score 1876;  DB 23;  Length 360;
Best Local Similarity   100.0%;  Pred. No. 2.1e-191;
Matches 360;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60

Qy     61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120

Qy    121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180

Qy    181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240

Qy    241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGI 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGI 300
```

```
Db    10  LFYTALLVFSALGNILALCLTCQKRRKINCTGIYLVHLAVSDLLFTVALPGRVVCVVLGS  69
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    95  SWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAI  154
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70  SWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPTVVCAHWGPCLRTAGRARLVCVAI  129
Qy   155  WTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSC  214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   130  WTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSC  189
Qy   215  YMKITWKLCS------------------------------------------------  224
              |||||||||
Db   190  YMKITWKLCSTAGRTQXPAGKDTTGGASPGGPSDQPEKTPPAGAAQEDPVTPRERTPPAG  249
Qy   225  -----------TAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMA       268
                          ||||||||||||||||||||||||||||||||||||||||||
Db   250  QPERTQXPAGKDATGGTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMA  309
Qy   269  RGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGILKLKGSSSS  328
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   310  RGMLHLPSCAERRAFLLSLQATVALMNMNCGIDPIIYFFASTHYRKWLLGIIKLKGSSSS  369
Qy   329  SSSSSS  334
              ||||||
Db   370  SSSSST  375
```

RESULT 3
AAR54080
ID   AAR54080 standard; Protein; 361 AA.
AC   AAR54080;
XX
DT   25-MAR-2003 (updated)
DT   03-FEB-1995 (first entry)
XX
DE   Epstein Barr virus induced (EBI-2) polypeptide.
XX
KW   Epstein Barr virus; EBV; induction; detection; diagnosis;
KW   lymphocytes; antigen; growth; differentiation; mediator;
KW   infectious mononucleosis.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Modified-site   8
FT   /note= "Potential N-linked glycosylation site."
FT   Region   32..57
FT   /label= Hydrophobic region.
FT   Region   78..95
FT   /label= Hydrophobic region.
FT   Region   106..127
FT   /label= Hydrophobic region.
FT   Region   126..134
FT   /note= "This sequences motif
FT   (S-II/V]-D-R-[Y/F]-X-X-X-X) is highly
FT   conserved among a large number of G-protein
FT   coupled receptors."
FT   Region   150..168
FT   /label= Hydrophobic region.
FT   Region   193..215
FT   /label= Hydrophobic region.
FT   Region   242..265
FT   /label= Hydrophobic region.
FT   Region   288..312
FT   /label= Hydrophobic region.
XX
PN   WO9412519-A1.
XX
PD   09-JUN-1994.
XX
PF   08-OCT-1993; 93WO-US09636.
XX

PR   25-NOV-1992; 92US-0980518.
XX
PA   (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX
PI   Birkenbach M, Kieff E;
XX
DR   WPI; 1994-200183/24.
DR   N-PSDB; AAQ64126.
XX
PT   DNA coding for Epstein Barr virus induced (EBI) polypeptide(s)
PT   and antibodies to EBI1, 2 and 3 - useful for detecting EBV by
PT   hybridisation or by immunoassay
XX
PS   Claim 8; Page 58-60; 84pp; English.
XX
CC   EBV infected B lymphocytes recapitulate features of antigen
CC   stimulation in enlarging, increasing RNA synthesis, expressing
CC   activation antigens and adhesion molecules, secreting Ig and
CC   proliferating. Unlike antigen stimulated B lymphocytes, EBV
CC   infected B lymphocytes continue to proliferate (in vitro) as
CC   immortalised lymphoblastoid cell lines. Because of the similar
CC   effects of EBV and antigen, EBV induced genes are likely to include
CC   mediators of antigen induced B lymphocyte growth or differentiation.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   361 AA;

Query Match   33.7%;   Score 632.5; DB 15; Length 361;
Best Local Similarity   37.8%; Pred. No. 1.1e-58;
Matches 135; Conservative 74; Mismatches 141; Indels   7; Gaps   4;

```
Qy     2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIPLRSLFYTALLVFSALGNILALCLTCQKSR   60
             ::  ||||: :              : ||::|||: ||||: ::|:||||||
Db     3  IQMANNFTPFSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK   61
Qy    61  KINCTGIYLVHLAVSDLLFTVALPGRVVCVVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL  120
             |||:||:|||:||||::||: :|:||::  |||  ||||||||||||||||:|||||
Db    62  KINSTTLYSTNLVISDILFTTALPTRIAYYAMGPDWRIGDALCRITALVFYINTYAGVNF  121
Qy   121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
             |  |:|||||::|||||||: :|:|||:|:|||||||| :: |||||||||||:|: :
Db   122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWWILVFAQTLPLLINPMSKQEAERITC  181
Qy   181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCIYMKITWKLCSTAREDPVTSRKGRHWR  240
             |||  |:||||::|:|||| || |:|: ||: :|: |::|:||
Db   182  MEYPNFEETKSLPWILLGACPIGVVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241
Qy   241  GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC  298
             | ||:  :||::|| ||  :|| |: :|       ::  :|:|  ||  | |||:|  |
Db   242  ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301
Qy   299  GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKAS---SETPSITQARGS  352
             |::|||||||||||||||||:||||||:||::|:  :| ::: :|:
Db   302  CMDPFIYFFACKGYKRKVVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS  358
```

RESULT 4
AAW53623
ID   AAW53623 standard; Protein; 361 AA.
XX
AC   AAW53623;
XX
DT   09-JUL-1998 (first entry)
XX
DE   Epstein Barr virus induced gene 2 (EBI-2).
XX
KW   Assessing; monitoring; foetal development; placental development;
KW   Epstein Barr virus; EBV; induced gene 2; EBI-2.
XX
OS   Homo sapiens.
XX
PN   US5744301-A.
XX

Left column:

```
PD   28-APR-1998.
XX
PF   02-FEB-1995;    95US-0383750.
XX
PR   02-FEB-1995;    95US-0383750.
PR   25-NOV-1992;    92US-0980518.
PR   30-NOV-1994;    94US-0352678.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
PI   Birkenbach M, Kieff E;
XX
DR   WPI; 1998-271060/24.
DR   N-PSDB; AAV25491.
XX
PT   Assessing or monitoring foetal or placental development - comprises
PT   detecting the level or size of Epstein Barr virus induced nucleic
PT   acid or protein in maternal serum samples
XX
PS   Example 2; Columns 43-46; 45pp; English.
XX
CC   The present sequence was used in the development of a novel method
CC   for assessing or monitoring foetal or placental development. The
CC   method comprises taking a maternal serum sample, and detecting the
CC   level or size of Epstein Barr virus (EBV) induced gene or protein 3
CC   (EBI-3) to obtain a result, which can be compared to a control to
CC   assess or monitor foetal or placental development.
XX
SQ   Sequence   361 AA;

Query Match           33.7%;  Score 632.5;  DB 19;  Length 361;
Best Local Similarity 37.8%;  Pred. No. 1.1e-58;
Matches  135;  Conservative 74;  Mismatches 141;  Indels 7;  Gaps 4;

Qy    2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
          ::   ||||    ||     |  |: |      |   :|::|||:|| |||:|| ||||    :
Db    3 IQMANNFTPPSATPQGNDCDLYAHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
         ||||   |   |||  |:||:|| || :|      ||||| :|||| |||:||   |   ::
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy  121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
         |::   | |||        :::   |   ::  |||  ||:|| |:||| ||:||  ||
Db  122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy  181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
         |:|:||| :  |||||  :| |    :| :|| |   ||| |  |: :|   |:|:   :|
Db  182 MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy  241 GCLLTLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNNC 298
          |:|||||||:||  :  |:  ||:|   ||  :    :|  ||||  |  :  |:|| |   |
Db  242 ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy  299 GITPIIYFFASTHYRKWLLGILKLGSSSSSSSSSSTPGKAS--SETPSITQARGS 352
          :  ||||||   :    |:    |:  |   :    :  |  : :
Db  302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS 358

RESULT 5
AAY90630
ID   AAY90630 standard; Protein; 361 AA.
XX
AC   AAY90630;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human G protein-coupled receptor EBI2.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist.
XX
```

Right column:

```
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;    99WO-US23938.
XX
PR   13-OCT-1998;    98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30635.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 1; Page 160-162; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY0643-
CC   AAY0677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence
CC   X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC   or Ala, and is preferably Lys. When the endogenous residue at this
CC   position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC   The 15 amino acid stretch between the substituted amino acid and the Pro
CC   may be endogenous, non-endogenous, or a mixture of endogenous and
CC   non-endogenous residues. The constitutively active GPCRs are useful for
CC   identifying antagonists, agonists and partial agonists for use as
CC   pharmaceutical agents. The mutant proteins are also useful in research
CC   settings for elucidating the roles of the receptors in normal and
CC   diseased conditions. Antagonists for a particular GPCR are useful for
CC   treating diseases and disorders associated with that receptor. Because
CC   the novel mutant GPCRs are constitutively active, they can be used
CC   directly for screening of compounds without the need for endogenous
CC   ligands. The present sequence represents a human wild-type GPCR referred
CC   to in an exemplification of the invention.
XX
SQ   Sequence   361 AA;

Query Match           33.7%;  Score 632.5;  DB 21;  Length 361;
Best Local Similarity 37.8%;  Pred. No. 1.1e-58;
Matches  135;  Conservative 74;  Mismatches 141;  Indels 7;  Gaps 4;

Qy    2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
          ::   ||||    ||     |  |: |      |   :|::|||:|| |||:|| ||||    :
Db    3 IQMANNFTPPSATPQGNDCDLYAHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
         ||||   |   |||  |:||:|| || :|      ||||| :|||| |||:||   |   ::
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy  121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
         |::   | |||        :::   |   ::  |||  ||:|| |:||| ||:||  ||
Db  122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy  181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
         |:|:||| :  |||||  :| |    :| :|| |   ||| |  |: :|   |:|:   :|
Db  182 MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy  241 GCLLTLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNNC 298
          |:|||||||:||  :  |:  ||:|   ||  :    :|  ||||  |  :  |:|| |   |
Db  242 ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301
```

Qy   299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKAS---SETPSTTQARGS 352
          . ||||| ||:::||| |||||| | | : |:| |||:::||:||
Db   302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMIHSKSS 358

RESULT 6
ABP81814
ID   ABP81814 standard; Protein; 361 AA.
XX
AC   ABP81814;
XX
DT   04-MAR-2003 (first entry)
XX
DE   Human EBV-induced gene 2 protein SEQ ID NO:112.
XX
KW   G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
KW   G protein-coupled receptor modulator; antibody; immune-related disease;
KW   growth-related disease; cell regeneration-related disease; AIDS; cancer;
KW   immunological-related cell proliferative disease; autoimmune disease;
KW   Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
KW   osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
KW   graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
KW   psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
KW   mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
KW   hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
KW   ulcer.
OS   Homo sapiens.
XX
PN   WO200261087-A2.
XX
PD   08-AUG-2002.
XX
PF   19-DEC-2001; 2001WO-US50107.
XX
PR   19-DEC-2000; 2000US-257144P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Burmer GC, Roush CL, Brown JP;
XX
DR   WPI; 2003-046718/04.
DR   N-PSDB; ABZ42660.
XX
PT   New isolated antigenic peptides e.g., for G protein-coupled receptors
PT   (GPCR), useful for diagnosing and designing drugs for treating
PT   conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,
PT   cancer or autoimmune diseases -
XX
PS   Disclosure; Fig 1; 523pp; English.
XX
CC   The present invention describes antigenic peptides (I) comprising:
CC   (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC   acids. Also described: (1) an assay for the detection of a particular
CC   G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC   and (2) an isolated antibody having high specificity and high affinity
CC   or avidity for a particular GPCR. (I) can be used as GPCR modulators and
CC   in gene therapy. The antigenic peptides for GPCRs are useful in detecting
CC   an antibody against a particular GPCR, and in the production of specific
CC   antibodies. The peptides and antibodies are also useful for detecting the
CC   presence or absence of corresponding GPCRs. The antigenic peptides for
CC   GPCRs and antibodies are useful for diagnosing and designing drugs for
CC   treating immune-related diseases, growth-related diseases, cell
CC   regeneration-related disease, immunological-related cell proliferative
CC   diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC   atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC   osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC   inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC   disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC   anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC   loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
CC   hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC   any other disorder in which GPCRs are involved. The antibodies may be

CC   used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode
CC   GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   361 AA;

Query Match   33.7%; Score 632.5; DB 24; Length 361;
Best Local Similarity   37.8%; Pred. No. 1.1e-58;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIILALCLTCQKSR 60
         |:: ||||: | |:||||| : | : :|||: : ::: ||| |:: ||| |
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK   61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRITAFVLYTDTYGGVYL 120
         |||| |||:: :|:| :|: ||||:| | |: | | |||:||||| | |::||| ||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy   121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCCVAIWTIVLLQTMPLLLMPMTKPLVGKLAC 180
         | |:: |||| | || | ::: | : :| :||: :|:|::|::: |::
Db   122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC 181

Qy   181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
         |: | :|:| :| | | | | || : |:|| || | | :|||| |
Db   182 MEYPNFEETKSLPWILLGACFIGYVLPLIIIIICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy   241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMMMNC 298
         ||| | || :|:| : ::|:||:| | | | | | || | |
Db   242 ALNTIILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy   299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSSTPGKAS---SETPSTTQARGS 352
         . ||||| ||:::||| |||||| | | : |:| |||:::||:||
Db   302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMIHSKSS 358

RESULT 7
AAY90664
ID   AAY90664 standard; Protein; 361 AA.
XX
AC   AAY90664;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human mutant G protein-coupled receptor EBI2 (L243K).
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist; mutein; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US23938.
XX
PR   13-OCT-1998;   98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30730.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 2; Page 262-263; 341pp; English.
XX

```
CC  The invention relates to constitutively active, non-endogenous versions
CC  of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC  AAY90677 and AAY90683-Y90687 and to DNA encoding them (AAA30709-A30743
CC  and AAA30775-A30779). The mutant proteins of the invention contain a
CC  mutation in a portion of the protein comprising intracellular loop 3
CC  (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC  is substituted for an endogenous proline in TM6 to form a sequence
CC  acids N-terminal of an endogenous amino acid is selected from Lys, His, Arg
CC  X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC  or Ala, and is preferably Lys. When the endogenous residue at this
CC  position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC  The 15 amino acid stretch between the substituted amino acid and the Pro
CC  may be endogenous, non-endogenous, or a mixture of endogenous and
CC  non-endogenous residues. The constitutively active GPCRs are useful for
CC  identifying antagonists, agonists and partial agonists for use as
CC  pharmaceutical agents. The mutant proteins are also useful in research
CC  settings for elucidating the roles of the receptors in normal and
CC  diseased conditions. Antagonists for a particular GPCR are useful for
CC  treating diseases and disorders associated with that receptor. Because
CC  the novel mutant GPCRs are constitutively active, they can be used
CC  directly for screening of compounds without the need for endogenous
CC  ligands. Sequences AAY90643- AAY90677 and AAY90683-Y90687 the mutant
CC  human GPCRs of the invention.
XX
SQ  Sequence   361 AA;

Query Match             33.6%;  Score 630.5;  DB 21;  Length 361;
Best Local Similarity   37.8%;  Pred. No. 1.8e-58;
Matches  135;  Conservative  74;  Mismatches 141;  Indels   7;  Gaps   4;

Qy      2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR  60
            ::   ||||    ||     :||:  |:  :  :: :||:|:||||:|||:||::||
Db      3  IQMANNFTPPSATPQGNDCDLYAHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK  61

Qy     61  KINCTGIYLVHLAVSDLLFTVALPGRVVCYLGSSWPFGKGLCRITAFVLYTDTYGGVVL 120
            ||| : |||:| ||: :|||:||  : : :   |||:|| ||  ||| :||::|   ||
Db     62  KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
            ||  | |: ||||  :||            : ||| |:|:|||||| |||||||||||||
Db    122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC 181

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
            ||| |:|||  |||:|||||||||||||||||:||||||||| ||  :  :  |  | :
Db    182  MEYPNFETKSLPWILLGACFIGYVLPLIIIIICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy    241  GCLLTLLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQCATVALMNMNC 298
            ||| :|| | :|| : |||:|||:::      |:||:| :||  :|   ||| ||| |||
Db    242  AKNTILIIVVFVLCFTPYHVALIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy    299  GITPIIYFFASTHYRKWLLGIILKLKGSSSSSSSSSTPGKAS---SETPSITQARGS 352
            ||||||||||    :| ||    ||| : :    :: :   ||||:: :: :: ||
Db    302  CMDPFIYFFACKGYGRKXVMRMLKRQVSVSISSAVKSAPEBNSREMTETOMMIHSKSS 358
```

RESULT 8
ABG66698
ABG66698
ID ABG66698 standard; Protein; 110 AA.
XX
AC ABG66698;
XX
DT 30-AUG-2002 (first entry)
XX
DE Human novel polypeptide #33.
XX
KW Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;

```
KW  fungal infection.
XX
OS  Homo sapiens.
XX
PN  WO20024340-A2.
XX
PD  06-JUN-2002.
XX
PF  30-NOV-2001; 2001WO-US47004.
XX
PR  30-NOV-2000; 2000US-0028952.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Tang YT, Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Wang J,  Wang D;
PI  Yamazaki V, Ujwal ML,  Drmanac RT;
XX
DR  WPI; 2002-508509/54.
DR  N-PSDB; ABK94922.
XX
PT  Novel nucleic acids and polypeptides for diagnosis, treatment of
PT  inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT  disorders, cancer and promoting wound healing -
XX
PS  Claim 10; Page 600; 672pp; English.
XX
CC  The invention relates to human novel polynucleotides and associated
CC  polypeptides. The polynucleotides and polypeptides are useful for
CC  treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC  disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC  and cancer and for promoting wound healing. The sequences are used to
CC  induce the proliferation of neural cells and regeneration of nerve and
CC  brain tissue, and are useful for the treatment of central and peripheral
CC  nervous system diseases and neuropathies, such as Alzheimer's disease,
CC  Parkinson's disease, Huntington's disease and amyotrophic lateral
CC  sclerosis. The sequences are involved in chemotactic or chemokinetic
CC  activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC  cell disorders and platelet disorders such as thrombocytopenia,
CC  regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC  growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC  osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC  disease. The sequences of the invention are also useful for gut
CC  protection or regeneration and treatment of lung or liver fibrosis,
CC  reperfusion injury in various tissues, immune deficiencies and disorders
CC  including severe combined immunodeficiency (SCID), bacterial or fungal
CC  infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC  gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC  and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC  novel polypeptides of the invention.
XX
SQ  Sequence   110 AA;

Query Match             29.6%;  Score 555;  DB 23;  Length 110;
Best Local Similarity   99.1%;  Pred. No. 4.3e-51;
Matches  105;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MACVSVDHYPAVVCAHWGPCLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  60

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA 226
            |||||||||||||||||||||||||||||||||||||||||||||
Db     61  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA 106
```

RESULT 9
AAU25588
AAU25588
ID AAU25588 standard; Protein; 177 AA.
XX
AC AAU25588;
XX
DT 18-DEC-2001 (first entry)
XX

Qy   161 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIIlSCYMKITW 220
          ||||||||||||||||||||||||||||||||||||| ||| | |: |: |
Db   68 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGAAPH----------GPGGL---CH----W 110

Qy   221 KLCST-----------AREDPVTSRKGRHWRG 241
          | :::   :|:|||| || ||
Db   111 LLWASGDHPVLLYEDHLEAVQHSWENPVTSGKGHHRRG 148

RESULT 10
AAY90652
ID   AAY90652 standard; Protein; 339 AA.
XX
AC   AAY90652;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human mutant G protein-coupled receptor GPR17 (V234K).
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist; mutant; mutein.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US23938.
XX
PR   13-OCT-1998;   98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30718.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 2; Page 233-234; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence
CC   X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC   or Ala, and is preferably Lys. When the endogenous residue at this
CC   position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC   The 15 amino acid stretch between the substituted amino acid and the Pro
CC   may be endogenous, non-endogenous, or a mixture of endogenous and
CC   non-endogenous residues. The constitutively active GPCRs are useful for
CC   identifying antagonists, agonists and partial agonists for use as
CC   pharmaceutical agents. The mutant proteins are also useful in research
CC   settings for elucidating the roles of the receptors in normal and
CC   diseased conditions. Antagonists for a particular GPCR are useful for
CC   treating diseases and disorders associated with that receptor. Because
CC   the novel mutant GPCRs are constitutively active, they can be used
CC   directly for screening of compounds without the need for endogenous
CC   ligands Sequences AAY90643- AAY90677 and AAY90683-Y90687 the mutant
CC   human GPCRs of the invention.
XX
SQ   Sequence   339 AA;

DE   Human G Protein-Coupled Receptor (GPCR) polypeptide #35.
XX
KW   Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW   attention deficit disorder; anxiety; depression; bipolar disorder;
KW   neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW   metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW   type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW   cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW   viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
KW   antidepressant; anorectic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200162797-A2.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-US05676.
XX
PR   23-FEB-2000; 2000US-0184247.
PR   23-FEB-2000; 2000US-0184303.
PR   23-FEB-2000; 2000US-0184304.
PR   23-FEB-2000; 2000US-0184305.
PR   23-FEB-2000; 2000US-0184397.
PR   02-MAR-2000; 2000US-0186457.
PR   03-MAR-2000; 2000US-0186810.
PR   09-MAR-2000; 2000US-0188064.
PR   13-MAR-2000; 2000US-0188880.
PR   03-APR-2000; 2000US-0194344.
PR   23-JUN-2000; 2000US-0213861.
PR   11-JUL-2000; 2000US-0217369.
PR   11-JUL-2000; 2000US-0217370.
PR   14-JUL-2000; 2000US-0218337.
PR   20-JUL-2000; 2000US-0218492.
XX
PA   (PHAA) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G, Wood LS, Parodi LA, Lind P;
XX
DR   WPI; 2001-570628/64.
DR   N-PSDB; AAS42840.
XX
PT   New isolated nucleic acid encoding a new G-protein coupled receptor
PT   polypeptide for detecting receptor modulators that can treat mental
PT   disorders, such as schizophrenia, anxiety, depression, or obesity -
XX
PS   Claim 35; Page 83; 279pp; English.
XX
CC   Sequences AAU25554-AAU25616 represent human G-protein coupled receptor
CC   (GPCR) polypeptides of the invention. The proteins and their associated
CC   DNA sequences can be used to identify compounds which bind to GPCR
CC   polypeptides and in screening for compounds that modulate GPCR activity.
CC   By screening a human subject for the presence of mutations in GPCR DNA, a
CC   GPCR-related disorder or a genetic predisposition can be diagnosed. The
CC   sequences can also be used for treatment and prevention of mental
CC   disorders such as schizophrenia, attention deficit disorder, anxiety,
CC   depression, dementia and bipolar disorder, neurological disorders such as
CC   Huntington's disease, Parkinson's disease and Tourette's syndrome,
CC   metabolic disorders such as obesity, anorexia and type 2 diabetes,
CC   cardiovascular disorders such as thrombosis, myocardial infarction,
CC   cardiomyopathy and atherosclerosis, viral infections caused by HIV and
CC   cancers.
XX
SQ   Sequence   177 AA;

Query Match   22.2%;   Score 417; DB 22; Length 177;
Best Local Similarity   58.9%; Pred. No. 4.3e-36;
Matches   93; Conservative   8; Mismatches   17;   Indels   40;   Gaps   5;

Qy   101 GLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPPRLRTAGRARLVCVAIWTLVLL 160
          |:|| ||||||||||||||||||||||||||||||||||||||
Db   14 GVCALHRHL------RGVYLMACVSVDHYPAVVCAHWGPCLRTAGRARLVCVAIWTLVLL 67

```
Query Match      21.1%;   Score 395.5; DB 21; Length 339;
Best Local Similarity  30.5%;  Pred. No. 2e-33;
Matches   95; Conservative   57; Mismatches 142; Indels   17; Gaps   5;

Qy   36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
         ||:|||:||:||||:||  :::|||||||:|:    : :|   ||:| ::|
Db   37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
         ||:||: ||||||||||:|:|::||:|| :||| |:|   ||| | :
Db   97 WPFGEIACRLITGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
         :|:|:|   |||    |:::       |||   :      HHALVSLAVAFTPFITTVTC
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS----HHALVSLAVAFTPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
         |:|  |  | ::   | : :|       |||:|:|:|  |   ::
Db  212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
         |  :|| :     :: |    ||||||:|||: :|  |   |||
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
         ::  |
Db  327 KTNESSLSAKS 337
```

```
RESULT 11
AAR53752
ID   AAR53752 standard; Protein; 339 AA.
XX
AC   AAR53752;
XX
DT   25-MAR-2003  (updated)
DT   07-FEB-1995  (first entry)
XX
DE   Seven transmembrane receptor (R12).
XX
KW   Primer; seven transmembrane receptor; receptor; amplification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   Domain          33..57
FT                   /label= Transmembrane domain.
FT   Domain          68..90
FT                   /label= Transmembrane domain.
FT   Domain          106..127
FT                   /label= Transmembrane domain.
FT   Domain          145..168
FT                   /label= Transmembrane domain.
FT   Domain          193..217
FT                   /label= Transmembrane domain.
FT   Domain          233..251
FT                   /label= Transmembrane domain.
FT   Domain          290..312
FT                   /label= Transmembrane domain.
XX
PN   WO9412635-A2.
XX
PD   09-JUN-1994.
XX
PF   17-NOV-1993;  93WO-US11153.
XX
PR   17-NOV-1992;  92US-0977452.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Godiska R,  Gray PW,  Schweickart VL;
XX
DR   WPI; 1994-200264/24.
```

```
DR   N-PSDB; AAQ66178.
XX
PT   DNA encoding seven transmembrane receptors - used to develop
PT   prods. for use as therapeutic or diagnostic agents for conditions
PT   involving the receptors.
XX
PS   Example 10; Page 81-82; 100pp; English.
XX
CC   PCR using two primers (AAQ66174, AAQ66175) was performed to amplify a
CC   partial sequence of the R20 seven transmembrane coding sequence
CC   which was later used as a probe for isolating the R20 genomic clone
CC   (AAQ66176) from a human placenta DNA library.  During the isolation
CC   of the R20 gene, two weakly hybridising sequences were identified
CC   which had significant homology to other seven transmembrane
CC   receptors.  The probe was used to screen a human genomic foetal
CC   liver DNA library, and while the R20 gene could not be identified in
CC   this library, several weakly hybridising clones were plaque
CC   purified, subcloned and sequenced.  The two clones were designated
CC   R2 (AAQ66177) and R12.  This is the sequence encoded by the R12 clone.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence  339 AA;
```

```
Query Match      20.9%;   Score 391.5; DB 15; Length 339;
Best Local Similarity  30.5%;  Pred. No. 5.4e-33;
Matches   95; Conservative   57; Mismatches 142; Indels   17; Gaps   5;

Qy   36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
         ||:|||:||:||||:||  :::|||||||:|:    : :|   ||:| ::|
Db   37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
         ||:||: ||||||||||:|:|::||:|| :||| |:|   ||| | :
Db   97 WPFGEIACRLITGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
         :|:|:|   |||    |:::       |||   :      HHALVSLAVAFTPFITTVTC
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS----HHALVSLAVAFTPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
         |:|  |  | ::   | : :|       |||:|:|:|  |   ::
Db  212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
         |  :|| :     :: |    ||||||:|||: :|  |   |||
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
         ::  |
Db  327 KTNESSLSAKS 337
```

```
RESULT 12
AAW07617
ID   AAW07617 standard; Protein; 339 AA.
XX
AC   AAW07617;
XX
DT   28-FEB-1997  (first entry)
XX
DE   Human G-protein thrombin-like receptor.
XX
KW   G-protein; thrombin; receptor; diagnosis; therapy; agonist;
KW   antagonist; haemophilia; wound healing; restenosis; angina;
KW   inflammation.
XX
OS   Homo sapiens.
XX
PN   WO9639438-A1.
XX
PD   12-DEC-1996.
XX
PF   06-JUN-1995;  95WO-US07180.
```

```
FT   Domain           33..57
FT                    /note= "Transmembrane domain 1"
FT   Domain           68..90
FT                    /note= "Transmembrane domain 2"
FT   Domain           106..127
FT                    /note= "Transmembrane domain 3"
FT   Domain           145..168
FT                    /note= "Transmembrane domain 4"
FT   Domain           193..217
FT                    /note= "Transmembrane domain 5"
FT   Domain           233..251
FT                    /note= "Transmembrane domain 6"
FT   Domain           290..312
FT                    /note= "Transmembrane domain 7"
XX
PN   US5759804-A.
XX
PD   02-JUN-1998.
XX
PF   93US-0153848;          93US-0153848.
XX
PR   17-NOV-1993;
PR   17-NOV-1992;           92US-0977452.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Godiska R,  Gray PW,  Schweickart VL;
XX
DR   WPI; 1998-332132/29.
DR   N-PSDB; AAV18356.
XX
PT   DNA encoding V28 seven transmembrane receptor polypeptide - useful
PT   for producing recombinant polypeptide and anti-V28 antibodies, and
PT   in screening assays for V28 agonists and antagonists
XX
PS   Example 10; Columns 87-90; 56pp; English.
XX
CC   The present sequence represents the human R12 protein which is encoded
CC   by the full length genomic R12 clone (AAV18356) which was isolated from
CC   a human genomic fetal liver DNA library.  The invention claims for the
CC   full length V28 genomic DNA sequence (AAV18343) isolated from a human
CC   placenta genomic library.  The V28 (AAW48722) and R12 proteins are seven
CC   transmembrane (7TM) receptors which are probably involved in signal
CC   transduction.  The invention also claims that cells transformed with V28
CC   DNA can be used to produce the recombinant polypeptide, to produce
CC   anti-V28 antibodies or in screening assays for V28 agonists or
CC   antagonists.  The antibodies, agonists and antagonists could then be
CC   used to modulate V28 receptor-ligand binding, for e.g. in immunological
CC   and/or inflammatory events in vivo.
XX
SQ   Sequence   339 AA;

Query Match             20.9%;  Score 391.5;  DB 19;  Length 339;
Best Local Similarity   30.5%;  Pred. No. 5.4e-33;
Matches   95;  Conservative  57;  Mismatches  142;  Indels  17;  Gaps   5;

Qy        36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVVLGSS 95
             ||||:|:||:|:  ||:||  :.|| | ::::.||:||| ||||:|:|.|:  |: ||:
Db        37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy        96 WPFGKGLCRLITAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
             |||:|| || | ::|:| |: | :|:||   |:|:| |: |:| :|   |:|| |:|: :
Db        97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy       156 TIVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFPAIGPCGPVGIILSC 214
             | |:|  : ||:  :|| |:|:|||   :: |:|:|| :|:|:|  ::| :   || :||
Db       157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy       215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
             ||: :|  ||   :||   ||  ::| |::| | | |:| :  :|:  | ::  |
Db       212 YLLIIRSLRQGLRVE----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy       275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
             | |::|: | |||:   |: : |  | :|||:|||     :|||    ||:|   ||||
Db       267 ASCATQRIALALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPSFEG 326

Qy       329 SSSSSSTPGKA 339
             :: ||:|
Db       327 KTNESSLSAKS 337
```

RESULT 13
AAW48733
```
ID   AAW48733 standard; Protein; 339 AA.
XX
AC   AAW48733;
XX
DT   25-SEP-1998 (first entry)
XX
DE   Human R12 seven transmembrane (7TM) receptor.
XX
KW   V28; placenta; seven transmembrane receptor; 7TM; signal transduction;
KW   immunology; inflammation; R12.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
```

(HUMA-) HUMAN GENOME SCI INC.

```
XX
PR   06-JUN-1995;          95WO-US07180.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Gocayne JD,  Li Y,  Ruben SM;
XX
DR   WPI; 1997-043073/04.
DR   N-PSDB; AAT44092.
XX
PT   Human G-protein thrombin receptor, HIBEB69 - useful to identify
PT   (ant)agonists, for treatment of angina, restenosis, wound healing
PT   etc.
XX
PS   Claim 1; Fig 1A-D; 58pp; English.
XX
CC   A novel human 7-transmembrane receptor (AAW07617) has been putatively
CC   identified as a G-protein thrombin receptor. Its amino acid
CC   sequence was deduced from a cDNA clone (AAT44092) discovered in a
CC   human infant brain cDNA library.  Recombinant receptor can be
CC   produced in host cells (e.g. E. coli, COS-7, Sf9) and used to
CC   raise antibodies or to screen for (ant)agonist cpds.  Agonists
CC   can be used to treat e.g. haemophilia or to promote wound
CC   healing; antagonists can be used to treat inflammation, angina,
CC   restenosis, stoke, etc.
XX
SQ   Sequence   339 AA;

Query Match             20.9%;  Score 391.5;  DB 18;  Length 339;
Best Local Similarity   30.5%;  Pred. No. 5.4e-33;
Matches   95;  Conservative  57;  Mismatches  142;  Indels  17;  Gaps   5;

Qy        36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVVLGSS 95
             ||||:|:||:|:  ||:||  :.|| | ::::.||:||| ||||:|:|.|:  |: ||:
Db        37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy        96 WPFGKGLCRLITAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
             |||:|| || | ::|:| |: | :|:||   |:|:| |: |:| :|   |:|| |:|: :
Db        97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy       156 TIVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFPAIGPCGPVGIILSC 214
             | |:|  : ||:  :|| |:|:|||   :: |:|:|| :|:|:|  ::| :   || :||
Db       157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy       215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
             ||: :|  ||   :||   ||  ::| |::| | | |:| :  :|:  | ::  |
Db       212 YLLIIRSLRQGLRVE----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy       275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
             | |::|: | |||:   |: : |  | :|||:|||     :|||    ||:|   ||||
Db       267 ASCATQRIALALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPSFEG 326

Qy       329 SSSSSSTPGKA 339
             :: ||:|
Db       327 KTNESSLSAKS 337
```

```
Db    267  ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326
```

```
Qy    329  SSSSSTPGKA 339
                :|:
Db    327  KTNESSLSAKS 337
```

RESULT 14
AAB21697
ID   AAB21697 standard; Protein; 339 AA.
XX
AC   AAB21697;
XX
DT   26-JAN-2001 (first entry)
XX
DE   Human 7TM receptor R12 cDNA clone protein.
XX
KW   Seven transmembrane receptor; 7TM; heptahelical; serpentine;
KW   G-protein-coupled; V28; V31; V12; R20; R2; R12; RM3; gene therapy;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   US6107475-A.
XX
PD   22-AUG-2000.
XX
PF   26-APR-1999;   99US-0299843.
XX
PR   17-MAY-1994;   94US-0245242.
PR   01-JUN-1998;   98US-0088337.
PR   17-NOV-1992;   92US-0977452.
PR   17-NOV-1993;   93US-0153848.
XX
PA   (ICOS-) ICOS CORP.
XX
PI   Schweickart VL, Gray PW, Godiska R;
XX
DR   WPI; 2000-571335/53.
DR   N-PSDB; AAA91725.
XX
PT   Polynucleotide encoding seven transmembrane receptors, antibody
PT   specific to the receptor, agonist and antagonist of the receptor useful
PT   for treating inflammation in a mammal -
XX
PS   Example 10; Columns 93-96; 61pp; English.
XX
CC   The present sequence is a novel seven transmembrane (7TM) receptors
CC   (also known as heptahelical, serpentine or G-protein-coupled receptors).
CC   The coding sequence for the present sequence may be used for gene
CC   therapy for diseases such as cancer.
XX
SQ   Sequence 339 AA;

Query Match   20.9%; Score 391.5; DB 21; Length 339;
Best Local Similarity   30.5%; Pred. No. 5.4e-33;
  Matches   95; Conservative   57; Mismatches 142; Indels   17; Gaps   5;

Qy   36 FYTALIVFSAIGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVLGSS 95
              |||||| ::|||||| ::|:|||:|| ||| :|||||
Db   37 FYILLDFILALVGNTLALMLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVVHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
              :||||:||| |:|| : ::| | :|||: | | :|||| ||:
Db   97 WPFGEIACRLTGFLPFYLANYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy   156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
              |: :|||: |||:|||:::   ::: |:|||||| ::|
Db   157 VVVAMAPLLLVSPQTVQTNHTVVCLQLYREKAS----HHALVSLAVAFTPPFITTVTC 211

Qy   215 YMKITWKLJCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCPSPYHLNIKCFMARGMLHL 274
              |:| ||| ::::|||| :|||| ||||||||| | :: ||: |::
Db   212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266
```

```
Qy    275  PSCAERRAFLLSLQATVALMMNCGITPIIYFFASTHYRKWLLGIL---KLKG--SSSS 328
               |||| ||||||||| ::|    :|||:||||:||||| ||||||:| :| :|||
Db    267  ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326
```

```
Qy    329  SSSSSTPGKA 339
                :|:
Db    327  KTNESSLSAKS 337
```

RESULT 15
AAY90618
ID   AAY90618 standard; Protein; 339 AA.
XX
AC   AAY90618;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human G protein-coupled receptor GPR17.
XX
KW   G protein-coupled receptor; GPCR; constitutively active;
KW   intracellular loop 3; transmembrane domain 6; drug screening;
KW   agonist; antagonist.
XX
OS   Homo sapiens.
XX
PN   WO200022129-A1.
XX
PD   20-APR-2000.
XX
PF   12-OCT-1999;   99WO-US23938.
XX
PR   13-OCT-1998;   98US-0170496.
XX
PA   (AREN-) ARENA PHARM INC.
XX
PI   Behan DP, Chalmers DT, Liaw CW;
XX
DR   WPI; 2000-329165/28.
DR   N-PSDB; AAA30601.
XX
PT   Non-endogenous constitutively activated human G protein-coupled
PT   receptors, useful for identifying agonists for use as pharmaceutical
PT   agents -
XX
PS   Example 1; Page 126-127; 341pp; English.
XX
CC   The invention relates to constitutively active, non-endogenous versions
CC   of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC   AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC   and AAA30775-A30779). The mutant proteins of the invention contain a
CC   mutation in a portion of the protein comprising intracellular loop 3
CC   (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC   is substituted for an endogenous residue in IC3 at a position 16 amino
CC   acids N-terminal of an endogenous proline in TM6 to form a sequence
CC   X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC   or Ala, and is preferably Lys. When the endogenous residue at this
CC   position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC   The 15 amino acid stretch between the substituted amino acid and the Pro
CC   may be endogenous, non-endogenous, or a mixture of endogenous and
CC   non-endogenous residues. The constitutively active GPCRs are useful for
CC   identifying antagonists, agonists and partial agonists for use as
CC   pharmaceutical agents. The mutant proteins are also useful in research
CC   settings for elucidating the roles of the receptors in normal and
CC   diseased conditions. Antagonists for a particular GPCR are useful for
CC   treating diseases and disorders associated with that receptor. Because
CC   the novel mutant GPCRs are constitutively active, they can be used
CC   directly for screening of compounds without the need for endogenous
CC   ligands. The present sequence represents a human wild-type GPCR referred
CC   to in an exemplification of the invention.
XX
SQ   Sequence 339 AA;

```
Query Match      20.9%;   Score 391.5;  DB 21;  Length 339;
Best Local Similarity   30.5%;  Pred. No. 5.4e-33;
Matches   95;  Conservative   57;  Mismatches 142;  Indels   17;  Gaps    5;

Qy     36  FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS   95
               ||  :  ||::||  |||:|:    . :     ::|:||||:|||    : ||:|:|:
Db     37  FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH   96

Qy     96  WPFGKGLCRLTAFVLYTDYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW   155
               |||| :  |||: | :: | :  |::|  :: |   |   || : |  :    | :
Db     97  WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLIYAHLACAFLW   156

Qy    156  TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC   214
               :|  |:     ||   |::| ::|   |::           ||:  :  | :|:::|
Db    157  VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITVTC   211

Qy    215  YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL   274
               |:|||||    |  : |  |  | :|  ::| ::|    :|| |||||||:  :
Db    212  YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG   266

Qy    275  PSCAERRAFLSLQATVALMNNCGITPIIYFPASTHYRKWLLGIL---KLKG--SSSS   328
               |||  : |:  || :|::| |::: :|:| ::: : || : :   |||    ||||
Db    267  ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG   326

Qy    329  SSSSSSTPGKA   339
               ::  :| || :
Db    327  KTNESSLSAKS   337
```

Search completed: February 19, 2004, 19:33:00
Job time : 45 secs

```
                OM protein - protein search, using sw model

                        Copyright (c)  1993 - 2004  Compugen Ltd.
                                    GenCore version 5.1.6

Run on:       February 19, 2004, 19:32:10 ; Search time 21 Seconds
                                (without alignments)
                        725.329 Million cell updates/sec

Title:         US-10-041-615-34
Perfect score: 1876
Sequence:      1 MIKLGPYFTPPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:      328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters:         328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Issued Patents AA:*
          1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
          4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	635.5	33.9	348	3	US-08-852-824-17	Sequence 17, Appl
2	632.5	33.7	361	1	US-08-383-750-4	Sequence 4, Appli
3	632.5	33.7	361	3	US-08-352-678-4	Sequence 4, Appli
4	632.5	33.7	361	4	US-09-536-954-4	Sequence 4, Appli
5	632.5	33.7	361	5	US-09-170-496D-78	Sequence 78, Appl
6	632.5	33.7	361	5	PCT-US93-09636-4	Sequence 4, Appli
7	630.5	33.6	361	4	US-09-170-496D-206	Sequence 206, App
8	395.5	21.1	339	1	US-09-170-496D-182	Sequence 182, App
9	391.5	20.9	339	2	US-08-153-848-44	Sequence 44, Appl
10	391.5	20.9	339	4	US-08-812-871-3	Sequence 3, Appli
11	391.5	20.9	339	4	US-09-299-843A-44	Sequence 44, Appl
12	391.5	20.9	339	4	US-09-088-137B-44	Sequence 44, Appl
13	391.5	20.9	339	4	US-09-170-496D-32	Sequence 32, Appl
14	391.5	20.9	339	5	PCT-US93-11153-44	Sequence 44, Appl
15	391.5	20.9	339	5	PCT-US95-07180-2	Sequence 2, Appli
16	353	18.8	357	5	PCT-US95-07180-3	Sequence 3, Appli
17	340	18.1	425	1	US-08-313-553-13	Sequence 13, Appl
18	340	18.1	425	3	US-08-767-993-13	Sequence 13, Appl
19	339.5	18.1	425	1	US-07-657-769B-69	Sequence 69, Appl
20	339.5	18.1	425	1	US-08-097-938-7	Sequence 7, Appli
21	339.5	18.1	425	1	US-07-789-184-220	Sequence 220, App
22	339.5	18.1	425	1	US-08-476-000-7	Sequence 7, Appli
23	339.5	18.1	425	1	US-08-475-263-220	Sequence 220, App
24	339.5	18.1	425	1	US-08-472-840-7	Sequence 7, Appli
25	339.5	18.1	425	1	US-08-485-886-220	Sequence 220, App
26	339.5	18.1	425	2	US-08-477-362-220	Sequence 220, App
27	339.5	18.1	425	2	US-08-477-134-220	Sequence 220, App
28	339.5	18.1	425	2	US-08-911-320A-3	Sequence 3, Appli
29	339.5	18.1	425	2	US-08-476-976-7	Sequence 7, Appli
30	339.5	18.1	425	2	US-08-742-440A-7	Sequence 57, Appl
31	339.5	18.1	425	2	US-08-560-098A-57	Sequence 220, App
32	339.5	18.1	425	3	US-08-473-489A-220	Sequence 220, App
33	339.5	18.1	425	3	US-08-474-410-7	Sequence 3, Appli
34	339.5	18.1	425	3	US-08-485-695-220	Sequence 220, App
35	339.5	18.1	425	3	US-09-217-101-3	Sequence 7, Appli
36	339.5	18.1	425	3	US-08-018-760-220	Sequence 56, Appl
37	339.5	18.1	425	3	US-08-486-673B-7	Sequence 380, App
38	331	17.6	328	3	US-08-513-974B-56	Sequence 56, Appl
39	331	17.6	328	3	US-08-513-974B-380	Sequence 2, Appli
40	331	17.6	328	3	US-09-461-436B-56	Sequence 2, Appli
41	324	17.3	328	4	US-08-459-046-2	Sequence 56, Appl
42	324	17.3	328	4	US-09-102-710B-2	Sequence 56, Appl
43	324	17.3	355	1	US-08-461-244-2	
44	324	17.3	355	4	US-09-045-583-56	
45	324	17.3	355	4	US-09-534-185-56	

ALIGNMENTS

```
RESULT 1
US-08-852-824-17
; Sequence 17, Application US/08852824C
; Patent No. 6060272
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Human G-Protein Coupled Receptors
; FILE REFERENCE: 1488.1220000
; CURRENT APPLICATION NUMBER: US/08/852,824C
; CURRENT FILING DATE: 1997-05-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 348
; TYPE: PRT
; ORGANISM: genomic
US-08-852-824-17

Query Match             33.9%;  Score 635.5;  DB 3;  Length 348;
Best Local Similarity   38.5%;  Pred. No. 3.6e-39;
Matches  132;  Conservative  69;  Mismatches 139;  Indels  3;  Gaps  2;

Qy      2 IKLGPYFTPPTKITKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRK 61
            ::  ::::|   :|  ||||||||| ||||:::::| :|||:|:  :|  ||    ::
Db      1 IQMANNFTPPSATPQNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRKK 59

Qy     62 INCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLM 121
            :  |||  :|||||:|||| :||||  |||| | | |||||||||| |||:|:||:||| 
Db     60 INSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNFM 119

Qy    122 ACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACM 181
           | ::| :|| :: :  |  :  :|  |:: ::|:|:| :|  |::|: | :||:||:||:
Db    120 TCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITCM 179

Qy    182 EYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRG 241
          ||:  ::|:|||:|| ::||:   ||:|| ||:|| | | :|  :::|  ||:| : : :
Db    180 EYPNFEETKSLPWILLGACFIGVLPLIIIKICYSQICCKLFRTAKQNPLTEKSGVNKKA 239

Qy    242 CLITLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNCG 299
          |::||:|:||:|  |||| ||:||| :| |   :|:|| : | ||||::|: | ::||| 
Db    240 LNTILLIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNCC 299

Qy    300 ITPIYFFASTHYRKWLLGILKLKGSSSSSSSSTPDGKASSE 342
          : ::|| || :|:| ::  ::: |::  ::: :: :| |::
Db    300 MDFFIYFPACKGYRKKVMRMLKRQVSVSISSAVKSAPENSRE 342

RESULT 2
US-08-383-750-4
; Sequence 4, Application US/08383750
```

; Patent No. 5744301
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliot
; TITLE OF INVENTION: Epstein Barr Virus Induced Genes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.,
; STREET: Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,750
; FILING DATE: Herewith
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel, L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0627.3300001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-383-750-4

Query Match   33.7%;   Score 632.5; DB 1; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR 60
        |:: ||||:. || : |||::|| |:|| : :
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYGGVVL 120
        |||:|| :.|:||| : ||:| || || |||| |:||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy 121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
        : |:| ::| | :|| ::|;
Db 122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC 181

Qy 181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
        || || |:|| || :.: || |:::.: :
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPLIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy 241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
        | |: | ||| || ::: | : |||:|
Db 242 ALNTIILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKAS---SETPSITQARGS 352
        ||||:||| |::| :| |||:||| | ||||
Db 302 CMDPFIYFFACKGYKRKVMRLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS 358

RESULT 3
US-08-352-678-4
; Sequence 4, Application US/08352678
; Patent No. 6043351
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott

; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/352,678
; FILING DATE: 30-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/980,518
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gates, Edward R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: B0801/7044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-352-678-4

Query Match   33.7%;   Score 632.5; DB 3; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR 60
        |:: ||||:. || : |||::|| |:|| : :
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVLVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYGGVVL 120
        |||:|| :.|:||| : ||:| || || |||| |:||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy 121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
        : |:| ::| | :|| ::|;
Db 122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC 181

Qy 181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
        || || |:|| || :.: || |:::.: :
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPLIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy 241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
        | |: | ||| || ::: | : |||:|
Db 242 ALNTIILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKAS---SETPSITQARGS 352
        ||||:||| |::| :| |||:||| | ||||
Db 302 CMDPFIYFFACKGYKRKVMRLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS 358

RESULT 4
US-09-536-954-4
; Sequence 4, Application US/09536954
; Patent No. 6500926
; GENERAL INFORMATION:
; APPLICANT: Birkenbach, Mark
; APPLICANT: Kieff, Elliott
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES

; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-78

Query Match   33.7%; Score 632.5; DB 4; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

```
Qy    2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR  60
          |::   ||||       ||      :  :  :|:|| :||:||:  |:|||||| ::|| :
Db    3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK  61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
        |||.| :||:.:||.|:||||| |: | ::|  : : |: | |||:||:|| | ||||  |
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy  121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
        | |:| |:: |||  .|| ||  :|:     ||:|::|:  | |||:|:| ||:  ::| |
Db  122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy  181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGILLSCYMKITWKLCSTAREDPVTSRKGRHWR 240
        ||| : |  ||:| :|| |  || | ||  :||.|:  || ||  |   ::| |:|| ::
Db  182 MEYPNFEETKSLPWILLGACFIGVYLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy  241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
          | |:::::|||:|||:|||| :| :|       |  :|:  ::  ::|  :||:| :|
Db  242 ALNTILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301

Qy  299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSTPGKAS--SETPSITQARGS      352
         : ||||||| : | ||::|||:|| |||::: |: | ||   |  ||  ||:|:|
Db  302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS    358
```

RESULT 6
PCT-US93-09636-4
; Sequence 4, Application PC/TUS9309636
; GENERAL INFORMATION:
;   APPLICANT: Birkenbach, Mark
;   APPLICANT: Kieff, Elliot
;   TITLE OF INVENTION: Epstein Barr Virus Induced Genes
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;   STREET: 1100 New York Avenue N.W., Suite 600
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US93/09636
;   FILING DATE: herewith
;   CLASSIFICATION:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 361 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear

; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
;   STREET: 600 Atlantic Avenue
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02210
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/536,954
;   FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/352,678
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Gates, Edward R.
;   REGISTRATION NUMBER: 31,616
;   REFERENCE/DOCKET NUMBER: B0801/7044
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-720-3500
;   TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 361 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-536-954-4

Query Match   33.7%; Score 632.5; DB 4; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

```
Qy    2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR  60
          |::   ||||       ||      :  :  :|:|| :||:||:  |:|||||| ::|| :
Db    3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK  61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
        |||.| :||:.:||.|:||||| |: | ::|  : : |: | |||:||:|| | ||||  |
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy  121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
        | |:| |:: |||  .|| ||  :|:     ||:|::|:  | |||:|:| ||:  ::| |
Db  122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy  181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGILLSCYMKITWKLCSTAREDPVTSRKGRHWR 240
        ||| : |  ||:| :|| |  || | ||  :||.|:  || ||  |   ::| |:|| ::
Db  182 MEYPNFEETKSLPWILLGACFIGVYLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy  241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
          | |:::::|||:|||:|||| :| :|       |  :|:  ::  ::|  :||:| :|
Db  242 ALNTILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301

Qy  299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSTPGKAS--SETPSITQARGS      352
         : ||||||| : | ||::|||:|| |||::: |: | ||   |  ||  ||:|:|
Db  302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS    358
```

RESULT 5
US-09-170-496D-78
; Sequence 78, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein

PCT-US93-09636-4

Query Match   33.7%; Score 632.5; DB 5; Length 361;
Best Local Similarity 37.8%; Pred. No. 6.2e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR 60
          |::| :||:||| |:::||| |||||||||::| :|:|:|
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
          ||||| | : || ||||| | :|:|||| ||||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy 121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVILQTMPLLLMPMTKPLVGKLAC 180
          ||||||||||||||||||||||| :| |::| ::|:||||||| :||:||||| ||||:
Db 122 MTCLSIDRFIAVVHPLRYNKIRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy 181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
          ||:| | | | :||||| : :|||||
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPIIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy 241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
          :: ||||| | :||: ||::: | : :||:   : : |
Db 242 ALNTIILIVVFVLCFTPYHVAIIQHMIKLRPSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSTPGKAS---SETPSITQARGS 352
          ::| |||||| : ||||| |||
Db 302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS 358

RESULT 7
US-09-170-496D-206
; Sequence 206, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 206
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-206

Query Match   33.6%; Score 630.5; DB 4; Length 361;
Best Local Similarity 37.8%; Pred. No. 8.6e-39;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR 60
          |::| :||:||| |:::||| |||||||||::| :|:|:|
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK 61

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
          ||||| | : || ||||| | :|:|||| ||||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy 121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVILQTMPLLLMPMTKPLVGKLAC 180
          ||||||||||||||||||||||| :| |::| ::|:||||||| :||:||||| ||||:
Db 122 MTCLSIDRFIAVVHPLRYNKIRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy 181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
          ||:| | | | :||||| : :|||||
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPIIIILICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy 241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298

Db 242 ALNTIILIVVFVLCFTPYHVAIIQHMIKLRPSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSTPGKAS---SETPSITQARGS 352

Db 302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS 358

RESULT 8
US-09-170-496D-182
; Sequence 182, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 182
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-182

Query Match   21.1%; Score 395.5; DB 4; Length 339;
Best Local Similarity 30.5%; Pred. No. 1.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

Qy   36 FYTALLVFSALGNIIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
          |::: ||||||||:||| :|||||
Db   37 FYILDFILALVGNTLAWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPPLRTAGRARLVCVAIW 155
          |||: |:|| |||| | ||| |::| |:||
Db   97 WPFGEIACRLTGFLFLYNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy 156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
          ::| |:|| :| | |:| ::
Db 157 VVVAVAMAPLLVSPQTVQTNMTVVCLQLYREKAS-----HHALVSLAVAFTPFITTVTC 211

Qy 215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
          ::| | | : ||:|
Db 212 YLLIIRSLRQGLRVEKRLKTKAKR----MIAIVLAIFLVCFVFYHVNRSVYVLHYRSHG 266

Qy 275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGII---KLKG--SSSS 328
          ||: |:: | :|:| |
Db 267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy 329 SSSSSSTPGKA 339

Db 327 KTNESSLSAKS 337

RESULT 9
US-08-153-848-44
; Sequence 44, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
;   ADDRESSEE: Bicknell
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago

STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-153-848-44

```
Query Match       20.9%;  Score 391.5;  DB 1;  Length 339;
Best Local Similarity  30.5%;  Pred. No. 2.1e-21;
Matches   95;  Conservative   57;  Mismatches 142;  Indels  17;  Gaps   5;

Qy  36 FYTALLVFSAIGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
          :  :||| ||| |      :||| |:||||:  :||:|: :
Db  37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVIVLPTRLIVHFSGNH 96

Qy  96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
       |||| : |||| |:: :  |:: |     :|  ::|     :::
Db  97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKIRRPLYAHLACAFLW 156

Qy 156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
        :|:   | |:||      |::| :|    :  :
Db 157 VVVAVAMAPLLVSPQTVQTNHTVCLQLYREKAS-----HHALVSLAVAFTPFITTVVTC 211

Qy 215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
        ||   |   :         :  :   : :||||:  ||:   :|:
Db 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy 275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
        |||  ||  |:  |      :|||||||||:| :::|    ::::| ::
Db 267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy 329 SSSSSSTPGKA 339
         :: |
Db 327 KTNESSLSAKS 337
```

RESULT 10
US-08-812-871-3
; Sequence 3, Application US/08812871
; Patent No. 5955303
; GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guegler, Karl
APPLICANT: Muzong Cheng
TITLE OF INVENTION: NOVEL HUMAN CHEMOKINE RECEPTOR-LIKE
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,871
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0237 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 992700
US-08-812-871-3

```
Query Match       20.9%;  Score 391.5;  DB 2;  Length 339;
Best Local Similarity  30.5%;  Pred. No. 2.1e-21;
Matches   95;  Conservative   57;  Mismatches 142;  Indels  17;  Gaps   5;

Qy  36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
          :  :||| ||| |      :||| |:||||:  :||:|: :
Db  37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVIVLPTRLIVHFSGNH 96

Qy  96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
       |||| : |||| |:: :  |:: |     :|  ::|     :::
Db  97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKIRRPLYAHLACAFlW 156

Qy 156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
        :|:   | |:||      |::| :|    :  :
Db 157 VVVAVAMAPLLVSPQTVQTNHTVCLQLYREKAS-----HHALVSLAVAFTPFITTVVTC 211

Qy 215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
        ||   |   :         :  :   : :||||:  ||:   :|:
Db 212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy 275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
        |||  ||  |:  |      :|||||||||:| :::|    ::::| ::
Db 267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy 329 SSSSSTPGKA 339
         :: |
Db 327 KTNESSLSAKS 337
```

RESULT 11
US-09-299-843A-44
; Sequence 44, Application US/09299843A
; Patent No. 6107475
; GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 6107475Sel Seven Transmembrane Receptors

```
US-09-088-337B-44
;  Sequence 44, Application US/09088337B
;  Patent No. 6348574
   GENERAL INFORMATION:
      APPLICANT: Godiska, Ronald
                 Gray, Patrick W.
                 Schweikart, Vicki L.
      TITLE OF INVENTION: No. 6348574el Seven Transmembrane Receptors
      NUMBER OF SEQUENCES: 66
      CORRESPONDENCE ADDRESS:
         ADDRESSE: Marshall, O'Toole, Gerstein, Murray &
                   Borun
         STREET: 6300 Sears Tower, 233 South Wacker Drive
         CITY: Chicago
         STATE: Illinois
         COUNTRY: USA
         ZIP: 60606
      COMPUTER READABLE FORM:
         MEDIUM TYPE: Floppy disk
         COMPUTER: IBM PC compatible
         OPERATING SYSTEM: PC-DOS/MS-DOS
         SOFTWARE: PatentIn Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
         APPLICATION NUMBER: US/09/088,337B
         FILING DATE: 01-Jun-1998
         CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
         APPLICATION NUMBER: US 08/153,848
         FILING DATE: 17-NOV-1993
         APPLICATION NUMBER: US 07/977,452
         FILING DATE: 17-NOV-1992
      ATTORNEY/AGENT INFORMATION:
         NAME: No. 6348574and, Greta E.
         REGISTRATION NUMBER: 35,302
         REFERENCE/DOCKET NUMBER: 31794
      TELECOMMUNICATION INFORMATION:
         TELEPHONE: (312) 474-6300
         TELEFAX: (312) 474-0448
         TELEX: 25 3856
   INFORMATION FOR SEQ ID NO: 44:
      SEQUENCE CHARACTERISTICS:
         LENGTH: 339 amino acids
         TYPE: amino acid
         TOPOLOGY: linear
      MOLECULE TYPE: protein
      SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-088-337B-44

Query Match        20.9%;    Score 391.5;  DB 4;  Length 339;
Best Local Similarity  30.5%;  Pred. No. 2.1e-21;  Indels 17;  Gaps 5;
Matches        95;  Conservative  57;  Mismatches 142;

Qy   36 FYTALLVPSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
         ||   :: |||||  ||   :  : :||| |:  :: |::|||:| : |:||:|:
Db   37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
         || : :||| ||: | :      : |: |  : |||  |   :
Db   97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
         :|| :|::  :  |||:: : : ||   :
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
         | ::   |     |:||   |||||  : |  :  ::|
Db  212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
         |:|   ||  :  |  |  |: ||  |||:||||::|||:|||||:       |   |  :
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
```

```
      NUMBER OF SEQUENCES: 66
      CORRESPONDENCE ADDRESS:
         ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
         ADDRESSE: Borun
         STREET: 6300 Sears Tower, 233 South Wacker Drive
         CITY: Chicago
         STATE: Illinois
         COUNTRY: USA
         ZIP: 60606
      COMPUTER READABLE FORM:
         MEDIUM TYPE: Floppy disk
         COMPUTER: IBM PC compatible
         OPERATING SYSTEM: PC-DOS/MS-DOS
         SOFTWARE: PatentIn Release #1.0, Version #1.25
      FILING DATE:
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
         APPLICATION NUMBER: US 09/088,337
         FILING DATE: 01-JUN-1998
      PRIOR APPLICATION DATA:
         APPLICATION NUMBER: US 08/153,848
         FILING DATE: 17-NOV-1993
      PRIOR APPLICATION DATA:
         APPLICATION NUMBER: US 07/977,452
         FILING DATE: 17-NOV-1992
      ATTORNEY/AGENT INFORMATION:
         NAME: Jill E. Uhl
         REGISTRATION NUMBER: 43,213
         REFERENCE/DOCKET NUMBER: 27866/32059B
      TELECOMMUNICATION INFORMATION:
         TELEPHONE: (312) 474-6300
         TELEFAX: (312) 474-0448
         TELEX:
   INFORMATION FOR SEQ ID NO: 44:
      SEQUENCE CHARACTERISTICS:
         LENGTH: 339 amino acids
         TYPE: amino acid
         TOPOLOGY: linear
      MOLECULE TYPE: protein
US-09-299-843A-44

Query Match        20.9%;    Score 391.5;  DB 3;  Length 339;
Best Local Similarity  30.5%;  Pred. No. 2.1e-21;
Matches        95;  Conservative  57;  Mismatches 142;  Indels 17;  Gaps 5;

Qy   36 FYTALLVPSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
         ||   :: |||||  ||   :  : :||| |:  :: |::|||:| : |:||:|:
Db   37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
         || : :||| ||: | :      : |: |  : |||  |   :
Db   97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
         :|| :|::  :  |||:: : : ||   :
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
         | ::   |     |:||   |||||  : |  :  ::|
Db  212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVVLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
         |:|   ||  :  |  |  |: ||  |||:||||::|||:|||||:       |   |  :
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
         ::  ||
Db  327 KTNESSLSAKS 337

RESULT 12
```

```
Db    327 KTNESSLSAKS 337
              :: :|| |:
```

RESULT 13
US-09-170-496D-32
; Sequence 32, Application US/09170496D
; Patent No. 655339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 655339-Endogenous, Constitutively Activated Human G Protein-
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-32

Query Match   20.9%; Score 391.5; DB 4; Length 339;
Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

```
Qy     36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
             :|:||||||    ||: ||:   :|:|||:|:|: : |:|:  : : :
Db     37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy     96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAEWGPRLRTAGRARLVCVAIW 155
             ||||| |:.| |||:  :|: :||: | |||    || |:
Db     97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy    156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
             |:|| :|::| :::| :|: |:   ||:   :||:
Db    157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy    215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
             |:    ::     :||    :|||  ||||  :
Db    212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy    275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
             ||||| |:|    |:| :  ||:|| |:|||   ||||::|  |:|||    |:|:
Db    267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy    329 SSSSSTPGKA 339
             :: :|| |:
Db    327 KTNESSLSAKS 337
```

RESULT 14
PCT-US93-11153-44
; Sequence 44, Application PCT/US9311153
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: Novel Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11153
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-11153-44

Query Match   20.9%; Score 391.5; DB 5; Length 339;
Best Local Similarity 30.5%; Pred. No. 2.1e-21;
Matches 95; Conservative 57; Mismatches 142; Indels 17; Gaps 5;

```
Qy     36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
             :|:||||||    ||: ||:   :|:|||:|:|: : |:|:  : : :
Db     37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy     96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
             ||||| |:.| |||:  :|: :||: | |||    || |:
Db     97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy    156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
             |:|| :|::| :::| :|: |:   ||:   :||:
Db    157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211

Qy    215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
             |:    ::     :||    :|||  ||||  :
Db    212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy    275 PSCAERRAFLLSLQATVALMMNNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
             ||||| |:|    |:| :  ||:|| |:|||   ||||::|  |:|||    |:|:
Db    267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy    329 SSSSSTPGKA 339
             :: :|| |:
Db    327 KTNESSLSAKS 337
```

RESULT 15
PCT-US95-07180-2
; Sequence 2, Application PC/TUS9507180
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; APPLICANT: GOCAYNE, JEANINE D
; APPLICANT: RUBEN, STEVEN M
; TITLE OF INVENTION: G-PROTEIN RECEPTOR HIBEB69
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRES:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; ADDRESSEE: STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NJ
; COUNTRY: US
; ZIP: 07068
; COMPUTER READABLE FORM:

```
         MEDIUM TYPE: Floppy disk
         COMPUTER: IBM PC compatible
         OPERATING SYSTEM: PC-DOS/MS-DOS
         SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
         APPLICATION NUMBER: PCT/US95/07180
         FILING DATE: 06-JUNE-1995
         CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
         NAME: MULLINS, J.G.
         REGISTRATION NUMBER: 30,073
         REFERENCE/DOCKET NUMBER: 325800-366
    TELECOMMUNICATION INFORMATION:
         TELEPHONE: 201-994-1700
         TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
         LENGTH: 339 amino acids
         TYPE: amino acid
         TOPOLOGY: linear
    MOLECULE TYPE: protein
PCT-US95-07180-2


Query Match           20.9%;  Score 391.5;  DB 5;  Length 339;
Best Local Similarity 30.5%;  Pred. No. 2.1e-21;
Matches   95; Conservative  57; Mismatches 142; Indels  17;  Gaps   5;

Qy     36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
             ||  | ||:|||:||| ||: :             :||:||||:|| | :    :|:
Db     37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVIVLFTRLVTHFSGNH 96

Qy     96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
             ||| |  ||||.| :|: : ::    : :||: :| |     : |   |  ||.|
Db     97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKQRRPLYAHLACAFLW 156

Qy    156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
             :|: :   |||:|  |:|  ::      :|| |  |:  ||:|:
Db    157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS----HHALVSLAVAFTFPFITTVTC 211

Qy    215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
             | :  ::|      .:|   :     ::||:||  |  |:|||| :::     |
Db    212 YLLIIRSLRQGLRVE----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy    275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL--KLKG--SSSS 328
             .||   |  :  :.|:.:    : .:|.|||||:::   |  :||   |||:     |||
Db    267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy    329 SSSSSTPGKA 339
             ::: || |:
Db    327 KTNESSLSAKS 337


Search completed: February 19, 2004, 19:35:29
Job time : 22 secs
```

OM protein - protein search, using sw model

;Copyright   GenCore version 5.1.6
(c) 1993 - 2004 Compugen Ltd.

Run on: February 19, 2004, 19:33:05 ; Search time 40 Seconds
(without alignments)
1884.441 Million cell updates/sec

Title: US-10-041-615-34
Perfect score: 1876
Sequence: 1 MIKLGPYFTPPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1876	100.0	360	12	US-10-041-615-34	Sequence 34, Appl
2	884	47.1	323	12	US-10-017-161-1620	Sequence 1620, Ap
3	884	47.1	323	12	US-10-292-798-1294	Sequence 1294, App
4	721	38.4	229	12	US-10-041-615-109	Sequence 109, App
5	635.5	33.9	348	9	US-09-827-937A-17	Sequence 17, Appl
6	632.5	33.7	361	12	US-10-041-615-110	Sequence 110, App
7	632.5	33.7	361	15	US-10-222-024-2	Sequence 2, Appli
8	632.5	33.7	361	15	US-10-251-385-78	Sequence 78, Appl
9	632.5	33.7	361	15	US-10-225-567A-112	Sequence 112, App
10	630.5	33.6	361	15	US-10-251-385-206	Sequence 206, App
11	417	22.2	177	11	US-09-791-932-95	Sequence 95, Appl
12	395.5	21.1	339	15	US-10-251-385-182	Sequence 182, App
13	391.5	20.9	339	9	US-09-848-889-12	Sequence 12, Appl
14	391.5	20.9	339	9	US-09-788-133-2	Sequence 2, Appli
15	391.5	20.9	339	10	US-09-828-478-4	Sequence 4, Appli
16	391.5	20.9	339	12	US-09-991-225-8	Sequence 8, Appli
17	391.5	20.9	339	12	US-10-349-021-4	Sequence 4, Appli
18	391.5	20.9	339	12	US-10-100-982-12	Sequence 12, Appl
19	391.5	20.9	339	12	US-10-268-332-14	Sequence 14, Appl
20	391.5	20.9	339	12	US-09-964-821B-14	Sequence 14, Appl
21	391.5	20.9	339	12	US-10-041-615-112	Sequence 112, App
22	391.5	20.9	339	15	US-10-251-385-32	Sequence 32, Appl
23	391.5	20.9	367	10	US-09-828-478-6	Sequence 6, Appli
24	391.5	20.9	367	12	US-10-149-021-6	Sequence 6, Appli
25	391.5	20.9	367	12	US-10-041-615-111	Sequence 111, App
26	391.5	20.9	367	15	US-10-225-567A-304	Sequence 304, App
27	384.5	20.5	298	12	US-10-041-615-113	Sequence 113, App
28	383	20.4	366	11	US-09-779-679-25	Sequence 25, Appl
29	377.5	20.1	177	12	US-09-801-944B-198	Sequence 198, App
30	346.5	18.5	330	12	US-10-348-190-21	Sequence 21, Appl
31	339.5	18.1	425	9	US-09-782-980-80	Sequence 80, Appl
32	339.5	18.1	425	10	US-09-684-430-4	Sequence 4, Appli
33	339.5	18.1	425	12	US-10-176-464A-3	Sequence 3, Appli
34	339.5	18.1	425	15	US-10-225-567A-324	Sequence 324, App
35	339.5	18.1	425	15	US-10-336-489-4	Sequence 4, Appli
36	339.5	18.1	425	15	US-10-177-293-130	Sequence 130, App
37	332	17.7	537	12	US-10-010-568-12	Sequence 12, Appl
38	332	17.7	537	15	US-10-311-956-4	Sequence 4, Appli
39	331	17.6	328	12	US-10-278-087A-56	Sequence 56, Appl
40	329	17.5	328	12	US-10-154-358-10	Sequence 10, Appl
41	329	17.5	328	12	US-09-745-842-18	Sequence 18, Appl
42	329	17.5	328	15	US-10-295-027-1288	Sequence 1288, Ap
43	329	17.5	328	15	US-10-225-567A-223	Sequence 223, App
44	328	17.5	355	12	US-09-826-509-479	Sequence 479, App
45	325	17.3	337	15	US-10-290-078-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-041-615-34
; Sequence 34, Application US/10041615
; Publication No.US20040014038A1
; GENERAL INFORMATION:
;   APPLICANT: Casman, Stacie J
;   APPLICANT: Edinger, Shlomit R
;   APPLICANT: Ellerman, Karen
;   APPLICANT: Smithson, Glennda
;   APPLICANT: Kekuda, Ramesh
;   APPLICANT: Padigaru, Muralidhara
;   TITLE OF INVENTION: No. US2004001403BAlel GPCR-Like Proteins and Nucleic Acids Encodi
;   FILE REFERENCE: 21402-233-061
;   CURRENT APPLICATION NUMBER: US/10/041,615
;   CURRENT FILING DATE: 2003-01-29
;   PRIOR APPLICATION NUMBER: 60/259,552
;   PRIOR FILING DATE: 2001-01-03
;   PRIOR APPLICATION NUMBER: 60/260,544
;   PRIOR FILING DATE: 2001-01-09
;   PRIOR APPLICATION NUMBER: 60/277,405
;   PRIOR FILING DATE: 2001-03-20
;   NUMBER OF SEQ ID NOS: 174
;   SOFTWARE: CuraSeqGist version 0.1
; SEQ ID NO 34
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-041-615-34

Query Match   100.0%; Score 1876; DB 12; Length 360;
Best Local Similarity   100.0%; Pred. No. 3.3e-164;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60

Qy   61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120

```
Db          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
     61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120

Qy   121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180

Qy   181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240

Qy   241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGI 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGI 300

Qy   301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
```

RESULT 2
US-10-017-161-1620
; Sequence 1620, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
;   APPLICANT: SUWA, MAKIKO
;   APPLICANT: ASAI, KIYOSHI
;   APPLICANT: AKIYAMA, YUTAKA
;   APPLICANT: ABURATANI, HIROYUKI
;   TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;   FILE REFERENCE: 084335/0152
;   CURRENT APPLICATION NUMBER: US/10/017,161
;   CURRENT FILING DATE: 2002-12-18
;   PRIOR APPLICATION NUMBER: JP 2001/246789
;   PRIOR FILING DATE: 2001-06-18
;   NUMBER OF SEQ ID NOS: 2430
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1620
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-017-161-1620

Query Match   47.1%;   Score 884; DB 12;   Length 323;
Best Local Similarity 65.7%;   Pred. No. 4.7e-73;
Matches 195; Conservative 8; Mismatches 36; Indels 58; Gaps 5;

```
Qy   101 GLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL 160
         |:|:|    |:|   ||||||||||||||||||||||||||||||||||||||||||||||
Db    48 GVCALHRHL-----RGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL 101

Qy   161 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITW 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITW 161

Qy   221 KLCS-----TAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFM--------- 267
         ||||      |||||||||||||     :   :      :
Db   162 KLCSTAGRTQTAREDPVTSGKGRHRRD-------------------SPGGPSDQQERTPLARLP 206

Qy   268 -------ARGMLHLPSCAERRAFLLSLQATVALMNMNCGIT--------------------PI 303
                        |     :   ::
Db   207 AYAADAGGRGGLLQPLPPQHQAVHGERDAPPAILCRAEGFLTVPSGHRGPHEHELWHYPI 266

Qy   304 IYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267 IYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 323
```

RESULT 3
US-10-292-798-1294
; Sequence 1294, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:

APPLICANT: SUWA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1294
LENGTH: 323
TYPE: PRT
ORGANISM: Homo sapiens
US-10-292-798-1294

Query Match   47.1%;   Score 884; DB 12;   Length 323;
Best Local Similarity 65.7%;   Pred. No. 4.7e-73;
Matches 195; Conservative 8; Mismatches 36; Indels 58; Gaps 5;

```
Qy   101 GLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL 160
         |:|:|    |:|   ||||||||||||||||||||||||||||||||||||||||||||||
Db    48 GVCALHRHL-----RGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL 101

Qy   161 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITW 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITW 161

Qy   221 KLCS-----TAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFM--------- 267
         ||||      |||||||||||||     :   :      :
Db   162 KLCSTAGRTQTAREDPVTSGKGRHRRD-------------------SPGGPSDQQERTPLARLP 206

Qy   268 -------ARGMLHLPSCAERRAFLLSLQATVALMNMNCGIT--------------------PI 303
                        |     :   ::
Db   207 AYAADAGGRGGLLQPLPPQHQAVHGERDAPPAILCRAEGFLTVPSGHRGPHEHELWHYPI 266

Qy   304 IYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   267 IYFFASTHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 323
```

RESULT 4
US-10-041-615-109
; Sequence 109, Application US/10041615
; Publication No. US2004014038A1
; GENERAL INFORMATION:
;   APPLICANT: Casman, Stacie J
;   APPLICANT: Edinger, Shlomit R
;   APPLICANT: Elleman, Karen
;   APPLICANT: Smithson, Glenda
;   APPLICANT: Kekuda, Ramesh
;   APPLICANT: Padigaru, Muralidhara
;   TITLE OF INVENTION: No. US2004001403841e1 GPCR-like Proteins and Nucleic Acids Encodi
;   FILE REFERENCE: 21402-233-061
;   CURRENT APPLICATION NUMBER: US/10/041,615
;   CURRENT FILING DATE: 2003-01-29
;   PRIOR APPLICATION NUMBER: 60/259,552
;   PRIOR FILING DATE: 2001-01-03
;   PRIOR APPLICATION NUMBER: 60/260,544
;   PRIOR FILING DATE: 2001-01-09
;   PRIOR APPLICATION NUMBER: 60/277,405
;   PRIOR FILING DATE: 2001-03-20
;   NUMBER OF SEQ ID NOS: 174
;   SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 109
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-041-615-109

Query Match 38.4%; Score 721; DB 12; Length 229;
Best Local Similarity 85.1%; Pred. No. 3e-58;
Matches 143; Conservative 5; Mismatches 12; Indels 8; Gaps 2;

```
Qy  59 SRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYTGGV 118
        ||||||||||||||||||||||||||||||||    :|: |   :
Db  66 SRKINCTGIYLVHLAVSDLLFTVALPGRVLAFRQGALQ--ADGVCALHRHL-----RGV 117

Qy 119 YLMACVSVDHYPAVVCAEHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKL 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118 YLMACVSVDHYPAVVCAEHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKL 177

Qy 179 ACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIIISCYMKITWKLCSTA 226
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 178 ACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIIISCYMKITWKLCSTA 225
```

RESULT 5
US-09-827-937A-17
; Sequence 17, Application US/09827937A
; Patent No. US2002005204 3A1
; GENERAL INFORMATION:
; APPLICANT: Li, Yi
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human G-Protein Coupled Receptors
; FILE REFERENCE: 1488.1220003
; CURRENT APPLICATION NUMBER: US/09/827,937A
; CURRENT FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 08/852,824
; PRIOR FILING DATE: 1997-05-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-937A-17

Query Match 33.9%; Score 635.5; DB 9; Length 348;
Best Local Similarity 38.5%; Pred. No. 3.6e-50;
Matches 132; Conservative 69; Mismatches 139; Indels 3; Gaps 2;

```
Qy   2 IKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRK 61
        ||:
Db   1 IQMANNFTPPSATPQNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVIVQNRKK 59

Qy  62 INCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYTGGVYLM 121
        ||:| ||:||||||||  :
Db  60 INSTTLYSTNLVISDLLFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNFM 119

Qy 122 ACVSVDHYPAVVCAEHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACM 181
        |:|:|:|||     :|: |:||| ||| :   :|| :
Db 120 TCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITCM 179

Qy 182 EYSSMESVLGLPLMVLVAFAIGFCGPVGIIISCYMKITWKLCSTAREDPVTSRKGRHWRG 241
        |:| |||||||||||||||||||||||||||  :
Db 180 EYPNFEETKSLPWILLGACFIGYVLPLIIIKICYSQICCKLFRTAKQNPLTEKSGVNKKA 239

Qy 242 CLLTLMLVAVVVCFSPYHLNLKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNCG 299
        :|::: ||| :  ||:|| |||| :
Db 240 LNTIILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNCC 299

Qy 300 ITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKASSE 342
        |:: :| |: :||:|||
Db 300 MDPFIYFFACKGYKRKVMRLKRQVSVSISSAVKSAPEENSRE 342
```

RESULT 6
US-10-041-615-110
; Sequence 110, Application US/10041615
; Publication No. US2004001403 8A1
; GENERAL INFORMATION:

; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glenda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US2004001403 8A1el GPCR-Like Proteins and Nucleic Acids Encod
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 110
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-110

Query Match 33.7%; Score 632.5; DB 12; Length 361;
Best Local Similarity 37.8%; Pred. No. 7.1e-50;
Matches 135; Conservative 74; Mismatches 141; Indels 7; Gaps 4;

```
Qy   2 IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
        ||:
Db   3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVIVQNRK 61

Qy  61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYTGGVYL 120
        ||:| ||:||||||||  :
Db  62 KINSTTLYSTNLVISDLLFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF 121

Qy 121 MACVSVDHYPAVVCAEHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
        |:|:|:|||     :|: |:||| ||| :   :|| :
Db 122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC 181

Qy 181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIIISCYMKITWKLCSTAREDPVTSRKGRHWR 240
        |:| |||||||||||||||||||||||||||  :
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPLIIIKICYSQICCKLFRTAKQNPLTEKSGVNKK 241

Qy 241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC 298
        :|::: ||| :  ||:|| |||| :
Db 242 ALNTIILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC 301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKAS---SETPSITCARGS 352
        |:: :| |: :||:|||
Db 302 CMDPFIYFFACKGYKRKVMRLKRQVSVSISSAVKSAPENSREMTETQMMIHSKSS 358
```

RESULT 7
US-10-222-024-2
; Sequence 2, Application US/10222024
; Publication No. US20030104487A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Ltd. (GB)
; APPLICANT: Pfizer Inc. (EP except GB, US, JP)
; TITLE OF INVENTION: Neuropeptide receptor and uses thereof
; FILE REFERENCE: PCS22032
; CURRENT APPLICATION NUMBER: US/10/222,024
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: GB 0119920.7
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-222-024-2

```
Query Match        33.7%;       Score 632.5; DB 15;   Length 361;
Best Local Similarity  37.8%;   Pred. No. 7.1e-50;
Matches 135;   Conservative  74;  Mismatches 141;   Indels  7;   Gaps    4;

Qy      2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR   60
           |::  ||||| |: : |  :| |:||:| ||:|| |::||||||
Db      3  IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK   61

Qy     61  KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL  120
           |||:  ::| |:||:: | :|| |
Db     62  KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF  121

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
           :::  |  |  || ||:|||
Db    122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC  181

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240
           |||                   |: :|   || |   |||||
Db    182  MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241

Qy    241  GCLLTLLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC  298
           |   |||| || ||:: |||||
Db    242  ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301

Qy    299  GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKAS---SETPSITQARGS  352
           |||||||| |:  |||
Db    302  CMDPFIYFFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS  358
```

RESULT 8
US-10-251-385-78
; Sequence 78, Application US/10251385
; Publication No. US2003010529241
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US2003010529241-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/10/251,385
;   CURRENT FILING DATE: 2002-09-20
;   PRIOR APPLICATION NUMBER: US/09/170,496
;   PRIOR FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-251-385-78

```
Query Match        33.7%;       Score 632.5; DB 15;   Length 361;
Best Local Similarity  37.8%;   Pred. No. 7.1e-50;
Matches 135;   Conservative  74;  Mismatches 141;   Indels  7;   Gaps    4;

Qy      2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR   60
           |::  ||||| |: : |  :| |:||:| ||:|| |::||||||
Db      3  IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK   61

Qy     61  KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL  120
           |||:  ::| |:||:: | :|| |
Db     62  KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF  121

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
           :::  |  |  || ||:|||
Db    122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC  181

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240
           |||                   |: :|   || |   |||||
Db    182  MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241
```

```
Qy    241  GCLLTLLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC  298
           |   |||| || ||:: |||||
Db    242  ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301

Qy    299  GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKAS---SETPSITQARGS  352
           |||||||| |:  |||
Db    302  CMDPFIYFFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS  358
```

RESULT 9
US-10-225-567A-112
; Sequence 112, Application US/10225567A
; Publication No. US2003011379841
; GENERAL INFORMATION:
;   APPLICANT: LifeSpan Biosciences
;   APPLICANT: Brown, Joseph P.
;   APPLICANT: Burmer, Glenna C.
;   APPLICANT: Roush, Christine L.
;   TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTOR
;   FILE REFERENCE: 1920-4-4
;   CURRENT APPLICATION NUMBER: US/10/225,567A
;   CURRENT FILING DATE: 2001-12-19
;   PRIOR APPLICATION NUMBER: 60/257,144
;   PRIOR FILING DATE: 2000-12-19
;   NUMBER OF SEQ ID NOS: 2292
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-225-567A-112

```
Query Match        33.7%;       Score 632.5; DB 15;   Length 361;
Best Local Similarity  37.8%;   Pred. No. 7.1e-50;
Matches 135;   Conservative  74;  Mismatches 141;   Indels  7;   Gaps    4;

Qy      2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR   60
           |::  ||||| |: : |  :| |:||:| ||:|| |::||||||
Db      3  IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK   61

Qy     61  KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL  120
           |||:  ::| |:||:: | :|| |
Db     62  KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF  121

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
           :::  |  |  || ||:|||
Db    122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVFAQTLPLLINPMSKQEAERITC  181

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240
           |||                   |: :|   || |   |||||
Db    182  MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241

Qy    241  GCLLTLLMLVAVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC  298
           |   |||| || ||:: |||||
Db    242  ALNTIILIIVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC  301

Qy    299  GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKAS---SETPSITQARGS  352
           |||||||| |:  |||
Db    302  CMDPFIYFFFACKGYKRKVMRMLKRQVSVSISSAVKSAPEENSREMTETQMMIHSKSS  358
```

RESULT 10
US-10-251-385-206
; Sequence 206, Application US/10251385
; Publication No. US2003010529241
; GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US2003010529241-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors

```
;  FILE REFERENCE: AREN-0040
;  CURRENT APPLICATION NUMBER: US/10/251,385
;  CURRENT FILING DATE: 2002-09-20
;  PRIOR APPLICATION NUMBER: US/09/170,496
;  PRIOR FILING DATE: 1998-10-13
;  NUMBER OF SEQ ID NOS: 294
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 206
;  LENGTH: 361
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-251-385-206

Query Match      33.6%;  Score 630.5; DB 15; Length 361;
Best Local Similarity    37.8%; Pred. No. 1.1e-49;
Matches 135; Conservative  74; Mismatches 141; Indels  7; Gaps  4;

Qy   2 IKLGPYFTDPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR  60
          |:::  |||||  ||  |    |       : |  : :|||||||||||||||||:  :
Db   3 IQMANNFFTPPSATPQGNDCLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVIVQNRK  61

Qy  61 KINCTGIYLVHLAVSDLLFTVALPGRVCVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
         ||| |   || ||||:|||| ||  || ||  :::   |  | |: ||  : |||:  ||
Db  62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWRIGDALCRITALVFYINTYAGVNF  121

Qy 121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
        | | ::|:|  ||   |  |  :| :    |||:|| :|| : |||||  |:| ||:
Db 122 MTCLSIDRFTAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC  181

Qy 181 MEYSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR   240
        |||  |   ::| |:|: ||| ||  |:|:|||: ||  |:  :|  |:|||||:|||:|
Db 182 MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241

Qy 241 GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC   298
          :|:| ::| :||: :|||  :| ::        ::|      |  :| |  ::|:| |
Db 242 AKNTILILIVFVLCFTPYHVALIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC   301

Qy 299 GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSTPGKAS---SETPSITQARGS       352
         : ||||||: :   : || :||||| ||||||||||   :   |:||: : |||
Db 302 CMDPIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPENSREMTETQMMIHSKSS      358


RESULT 11
US-09-791-932-95
;  Sequence 95, Application US/09791932
;  Publication No. US20030003451A1
;  GENERAL INFORMATION:
;   APPLICANT: Vogeli, Gabriel
;   APPLICANT: Parodi, Luis A.
;   APPLICANT: Hiebsch, Ronald R.
;   APPLICANT: Lind, Peter
;   APPLICANT: Kaytes, Paul S.
;   APPLICANT: Ruff, Valerie
;   APPLICANT: Huff, Rita M.
;   APPLICANT: Wood, Linda S.
;   TITLE OF INVENTION: No. US20030003451Ale1 G Protein-Coupled Receptors Cross-Referen
;   FILE REFERENCE: 00325.US1
;   CURRENT APPLICATION NUMBER: US/09/791,932
;   CURRENT FILING DATE: 2001-02-23
;   PRIOR APPLICATION NUMBER: 60/184,305
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,304
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,303
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,397
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,247
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/188,880
;   PRIOR FILING DATE: 2000-03-13
;   PRIOR APPLICATION NUMBER: 60/217,369
;   PRIOR FILING DATE: 2000-07-11
```

```
;   PRIOR APPLICATION NUMBER: 60/217,370
;   PRIOR FILING DATE: 2000-07-11
;   PRIOR APPLICATION NUMBER: 60/218,492
;   PRIOR FILING DATE: 2000-07-20
;   PRIOR APPLICATION NUMBER: 60/186,810
;   PRIOR FILING DATE: 2000-03-03
;   PRIOR APPLICATION NUMBER: 60/188,064
;   PRIOR FILING DATE: 2000-03-09
;   PRIOR APPLICATION NUMBER: 60/186,457
;   PRIOR FILING DATE: 2000-03-02
;   PRIOR APPLICATION NUMBER: 60/213,861
;   PRIOR FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: 60/194,344
;   PRIOR FILING DATE: 2000-04-03
;   PRIOR APPLICATION NUMBER: 60/218,337
;   PRIOR FILING DATE: 2000-07-14
;   NUMBER OF SEQ ID NOS: 184
;   SOFTWARE: PatentIn version 3.0
;   SEQ ID NO 95
;   LENGTH: 177
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-791-932-95

Query Match      22.2%;  Score 417; DB 11; Length 177;
Best Local Similarity    58.9%; Pred. No. 2e-30;
Matches  93; Conservative  8; Mismatches 17; Indels 40; Gaps  5;

Qy 101 GLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLL 160
           | | :|:                 ||||||||||||||||:|||||||||||||||||
Db  14 GVCALHRHL------RGVYLMACVSVDHYPAVVCAEHGPCLRTAGRARLVCVAIWTLVLL  67

Qy 161 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITW 220
        |||||||||||||||||||||||||||||||||
Db  68 QTMPLLLMPMTKPLVGKLACMEYSSMESVLGAAPH----------GPGGL---CH----W 110

Qy 221 KLCST----------------AREDPVTSRKGRHWRG                        241
                            : :|  |  |:||| |:||
Db 111 LLWASGDHPVLLYEDHLEAVQHSWENPVTSGKGHHRRG                       148


RESULT 12
US-10-251-385-182
;  Sequence 182, Application US/10251385
;  Publication No. US2003010529A1
;  GENERAL INFORMATION:
;   APPLICANT: Behan, Dominic P.
;   APPLICANT: Chalmers, Derek T.
;   APPLICANT: Liaw, Chen W.
;   TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
;   TITLE OF INVENTION: Protein-Coupled
;   TITLE OF INVENTION: Receptors
;   FILE REFERENCE: AREN-0040
;   CURRENT APPLICATION NUMBER: US/10/251,385
;   CURRENT FILING DATE: 2002-09-20
;   PRIOR APPLICATION NUMBER: US/09/170,496
;   PRIOR FILING DATE: 1998-10-13
;   NUMBER OF SEQ ID NOS: 294
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 182
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-251-385-182

Query Match      21.1%;  Score 395.5; DB 15; Length 339;
Best Local Similarity    30.5%; Pred. No. 4.1e-28;
Matches  95; Conservative  57; Mismatches 142; Indels 17; Gaps  5;

Qy  36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS  95
        ||    : |::|| ||| |||      :  :       |||:|  |  | ::   |:|::
Db  37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFMHLAVADLSCVLVLPTRLVYHFSGNH   96
```

Left column

```
Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
        |||:: |||| |: ::: ::: :|:|||||: :|:|
Db   97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
        |: :|||||| |||||||  ::    |||||||| ::
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTPPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
        : || |||||| ||| :: :|
Db  212 YLLIIRSLRQGLRVEKRLKTKAKR------MAIVLAIFLVCFVPYHVNRSVYLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
        |||:: :| ||| :|: ::| |||| : :|:|
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
        :|:
Db  327 KTNESSLSAKS 337
```

```
RESULT 13
US-09-848-889-12
;  Sequence 12, Application US/09848889
;  Patent No. US2002002555A1
;  GENERAL INFORMATION:
;  APPLICANT: Au-Young, Janice
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Cheng, Muzong
;  TITLE OF INVENTION: GPCR DIAGNOSTIC FOR BRAIN CANCER
;  FILE REFERENCE: PC-0042 CIP
;  CURRENT APPLICATION NUMBER: US/09/848,889
;  CURRENT FILING DATE:  2001-05-03
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PERL Program
;  SEQ ID NO 12
;  LENGTH: 339
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  OTHER INFORMATION: Incyte ID No. US2002002555A1 g992700
US-09-848-889-12

Query Match             20.9%;   Score 391.5;  DB 9;   Length 339;
Best Local Similarity   30.5%;   Pred. No. 9.5e-28;
Matches    95;  Conservative  57;  Mismatches 142;  Indels  17;  Gaps    5;
```

Right column

```
RESULT 14
US-09-788-133-2
;  Sequence 2, Application US/09788133
;  Patent No. US2002052001A1
;  GENERAL INFORMATION:
;  APPLICANT: GLAXO GROUP LTD
;  TITLE OF INVENTION: ASSAY
;  FILE REFERENCE: P79011
;  CURRENT APPLICATION NUMBER: US/09/788,133
;  CURRENT FILING DATE:  2001-02-20
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
;  SEQ ID NO 2
;  LENGTH: 339
;  TYPE: PRT
;  ORGANISM: homo sapiens
US-09-788-133-2

Query Match             20.9%;   Score 391.5;  DB 9;   Length 339;
Best Local Similarity   30.5%;   Pred. No. 9.5e-28;
Matches    95;  Conservative  57;  Mismatches 142;  Indels  17;  Gaps    5;
```

```
Qy   36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
        || ||||||||| | ||::|||| : :|:|||||||
Db   37 FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
        |||:: |||| |: ::: ::: :|:|||||: :|:|
Db   97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC 214
        |: :|||||| |||||||  ::    |||||||| ::
Db  157 VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS-----HHALVSLAVAFTPPFITTVTC 211

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
        : || |||||| ||| :: :|
Db  212 YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS 328
        |||:: :| ||| :|: ::| |||| : :|:|
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
        :|:
Db  327 KTNESSLSAKS 337
```

```
RESULT 15
US-09-828-478-4
;  Sequence 4, Application US/09828478
;  Patent No. US2002015528A1
;  GENERAL INFORMATION:
;  APPLICANT: Xiao, Yonghong
;  TITLE OF INVENTION: Regulation of Human CysLT2-Like GPCR
;  TITLE OF INVENTION: Protein
;  FILE REFERENCE: 04974.00458
;  CURRENT APPLICATION NUMBER: US/09/828,478
;  CURRENT FILING DATE:  2001-04-09
;  PRIOR APPLICATION NUMBER: 60/195,196
;  PRIOR APPLICATION FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: 60/254,876
;  PRIOR APPLICATION FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 4
;  LENGTH: 339
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-09-828-478-4

Query Match             20.9%;   Score 391.5;  DB 10;   Length 339;
Best Local Similarity   30.5%;   Pred. No. 9.5e-28;
Matches    95;  Conservative  57;  Mismatches 142;  Indels  17;  Gaps    5;
```

```
Qy    36  FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS  95
              |: .:|||||| :    .:  ::|:||||:|:|   : :||:|: .|:
Db    37  FYLLDFILALVGNTLALWLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH  96

Qy    96  WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW  155
              ||||: |||||   :.  :|:|||  :.:|::: :|:|          | ||:|:
Db    97  WPFGEIACRLTGFLFYLNMYASIYFLTCISADRPLAIVHPVKSLKIRRPLYAHLACAFLW  156

Qy   156  TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC  214
              ||:||  |||||:          | :. :|:|  |           |||: :::|
Db   157  VVVAVAMAPLLVSPQTVQTNHTVVCLQLYREKAS----HHALVSLAVAFTPPFITTVTC  211

Qy   215  YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL  274
              |: |:   :| :    .|:  .  : :|||  |||||:   :|||:     ::
Db   212  YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG  266

Qy   275  PSCAERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG---SSSS  328
              ||| ::|.    :| :::|||||||| ::||:|          :|   :||.   ||||
Db   267  ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKFRHALCNLLCGKRLKGPPPSFEG  326

Qy   329  SSSSSSTPGKA  339
              :: ||||
Db   327  KTNESSLSAKS  337
```

Search completed: February 19, 2004, 19:36:22
Job time : 41 secs

```
                Copyright   GenCore version 5.1.6
                          (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February 19, 2004, 19:30:15 ; Search time 20 Seconds
                             (without alignments)
                             1731.036 Million cell updates/sec

Title:       US-10-041-615-34
Perfect score: 1876
Sequence:    1 MIKLGPYFTPPTKIKTKIKD........SETPSITQARGSMFLAEHVV 360

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:       283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR.76:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	632.5	33.7	361	2	B45680	G protein-coupled
2	364.5	19.4	427	2	S17148	alpha-thrombin rec
3	339.5	18.1	425	2	A37912	thrombin receptor
4	339	17.8	432	2	A43448	thrombin receptor
5	332.5	17.7	420	2	I51667	thrombin receptor
6	329	17.5	328	2	JC4800	P2Y6 receptor - hu
7	324	17.3	355	2	JC5067	G protein-coupled
8	322.5	17.2	360	2	A57160	chemokine (C-C) re
9	315.5	16.8	371	2	JC5498	G protein-coupled
10	308	16.4	328	2	I55450	G protein-coupled
11	308	16.4	373	2	A47556	ATP receptor P2u -
12	307.5	16.4	362	2	A39714	G protein-coupled
13	307	16.3	355	2	A45177	chemokine (C-C) re
14	306.5	16.3	360	2	A53611	interleukin-8 rece
15	302	16.1	380	2	I38435	angiotensin recept
16	300.5	16.0	371	2	JC5796	probable chemoattr
17	300	16.0	367	2	JE0349	interferon-inducib
18	298.5	15.9	358	2	A53752	interleukin-8 rece
19	298.5	15.9	359	2	A48921	interleukin-8 rece
20	297.5	15.9	308	2	I50241	G protein-coupled
21	297	15.8	355	2	JQ1231	interleukin-8 rece
22	296	15.8	370	2	JC5549	heptahelical P2Ys-
23	293.5	15.6	362	2	A30341	G protein-coupled
24	293	15.6	395	2	S40685	probable G protein
25	291.5	15.5	344	2	T09508	intron 17 purinerg
26	290.5	15.5	375	2	A54946	P-2U nucleotide re
27	290.5	15.5	428	2	S30508	probable G protein
28	289.5	15.4	359	2	I49341	MIP-1 alpha recept
29	287	15.3	350	2	A39445	interleukin-8 rece
30	286.5	15.3	360	2	JC4587	chemokine (C-C) re
31	285	15.2	355	2	G02436	chemokine (C-C) re
32	284	15.1	354	2	I58186	probable G protein
33	284	15.1	428	2	A44021	somatostatin recep
34	283	15.1	372	2	I38532	delta opioid recep
35	282	15.0	356	2	I49340	MIP-1 alpha recept
36	281	15.0	373	2	JC4737	G protein-coupled
37	279.5	14.9	362	2	B57641	G protein-coupled
38	279.5	14.9	404	2	JC5784	adrenomedullin rec
39	278.5	14.8	369	2	JC2083	somatostatin recep
40	278.5	14.8	373	2	JC4162	P2Y receptor - bov
41	278	14.8	397	2	S66518	proteinase-activat
42	277.5	14.8	359	2	S44425	angiotensin II rec
43	275.5	14.7	362	2	S68207	G protein-coupled
44	275.5	14.7	398	2	A57510	mu opioid receptor
45	274.5	14.6	354	2	I53033	G protein-coupled

ALIGNMENTS

RESULT 1
B45680
G protein-coupled peptide receptor EBI 2 - human
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: B45680
R;Birkenbach, M.; Josefsen, K.; Yalamanchili, R.; Lenoir, G.; Kieff, E.
J. Virol. 67, 2209-2220, 1993
A;Title: Epstein-Barr virus-induced genes: first lymphocyte-specific G protein-coupled p
A;Reference number: A45680; MUID:93188173; PMID:8383238
A;Accession: B45680
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-361 <BIR>
A;Cross-references: GB:L08177; NID:g292056; PIDN:AAA35924.1; PID:g292057
A;Experimental source: B-lymphocytes
A;Note: sequence extracted from NCBI backbone (NCBIN:127096, NCBIP:127097)
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; transmembrane protein

```
Query Match             33.7%;  Score 632.5;  DB 2;  Length 361;
Best Local Similarity   37.8%;  Pred. No. 5.5e-48;
Matches  135;  Conservative  74;  Mismatches 141;  Indels  7;  Gaps  4;

Qy      2  IKLGPYFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSR   60
           |::.|||||.:| |:. :   ::||.|.||:.: :||.||||||||.|.:|| .:
Db      3  IQMANNFTPFSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVIVQNRK   61

Qy     61  KINCTGIYLVHLAVSDLLFTVALPGRVCVCVLGSSWPFGKGLCRITAFVLYTDTYGGVYL  120
           |:: ||:|:.:..:||||.||||.|.|.:|| ..|.||..||| |||:|.|| :.||:|
Db     62  KINSTTIYSTNLVISDILFTTALPTRIAYTAMGFDWRIGDALCRITALVFYINTYAGVNF  121

Qy    121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIKTLVLLQTMPLLLMPMTKPLVGKLAC  180
           ||||.|:||:|.|||:.|| |.:..::|.|..||:|:|.||.:||.|..|:.|.||::|
Db    122  MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFVWILVFAQTLPLLINPMSKQEAERITC  181

Qy    181  MEYSSMESVLGLPLMVLVAFAIGFCCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240
           ||| ::|| ||| |::|.|.: :||||||| |:: |:||.|:|| .|:|.|||. ||.:||
Db    182  MEYPNPBEETKSLPWILLGACFIGYVLPLIILICYSQICCKLFRTAKQNPLTEKSGVNKK  241

Qy    241  GCLLTLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC   298
           ||.|.|:.: ...|||::.|:::|.:|.| |:||| .:.|..: ||| :.|:||.:|:|
Db    242  ALNTILIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC   301

Qy    299  GITPIYIYFFASTHYRKWLLGILKLKGSSSSSSSSTPGKAS--SETPSITQARGS      352
           ||| ||:||||.:.:.|.::|:..:.|..: ::.:.  .|.  ::|...:|.:::
Db    302  CMDPFIYFFACKGYKRKVMRVMLRKQVSVSISSAVKSAPEENSREMTETQMMIHSKSS  358
```

RESULT 2
S17148
alpha-thrombin receptor - Chinese hamster

```
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S17148
R;Rasmussen, U.B.; Vouret-Craviari, V.; Jallat, S.; Schlesinger, Y.; Pages, G.; Pavirani
  FEBS Lett. 288, 123-128, 1991
A;Title: cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca
A;Reference number: S17148; MUID:91348847; PMID:1652467
A;Accession: S17148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <RAS>
A;Cross-references: EMBL:X61958; NID:g940495; PIDN:CAA43957.1; PID:g49538
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match             19.4%;  Score 364.5;  DB 2;  Length 427;
Best Local Similarity   30.7%;  Pred. No. 2.3e-24;
Matches   96;  Conservative   58;  Mismatches 146;  Indels  13;  Gaps   6;

Qy    25 SQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP 84
Db   100 SPWLRLFIPSV-YTFVFVVSLPLNILAIAVFVLKMKVKKPAVVYMLHLAMADVLFVSVLP 158

Qy    85 GRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTA 144
Db   159 LKISYYFSGSGDWQFGSELCRFVTAAFYCNMYASIMLMTVISIDRFLAVVYPIQSLSWRTL 218

Qy   145 GRARLVCVAIWTLVLLQTMPLLLMPMTK--PLVGKLACMEYSSMESVLGLPLMVLVAF-A 201
Db   219 GRANFTCLVIWVMAIMGVVPLLLKEQTRVPGLNITTCHDVLNETLLQGFYSYRFSAFSA 278

Qy   202 IGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHL 261
Db   279 VPFLVPLIISTICYMSIIRCLSSSS----VANRSKKSRALFLSAAVFCVFIVCFGP--T 331

Qy   262 NIKQFMARGMLHLPSCAERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGILK 321
Db   332 NVLLIMHYLLLSDSPATEKAYFAYLL--CVCVSSVSCCIDPLIYYASSECQRTLYGILC 389

Qy   322 LKGSSSSSSSS 334
Db   390 CKESSDPNSYNST 402


RESULT 3
A37912
thrombin receptor precursor - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 05-Nov-1999
C;Accession: A37912
R;Vu, T.K.H.; Hung, D.T.; Wheaton, V.I.; Coughlin, S.R.
  Cell 64, 1057-1068, 1991
A;Title: Molecular cloning of a functional thrombin receptor reveals a novel proteolytic
A;Reference number: A37912; MUID:91168254; PMID:1672265
A;Accession: A37912
A;Molecule type: mRNA
A;Residues: 1-425 <VUA>
A;Cross-references: GB:M62424; NID:g339676; PIDN:AAA36743.1; PID:g339677
C;Genetics:
C;Gene: F2R
A;Gene: GDB:F2R
A;Cross-references: GDB:127737; OMIM:187930
A;Map position: 5q13-5q13
C;Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-425/Product: thrombin receptor #status predicted <MAT>

Query Match             18.1%;  Score 339.5;  DB 2;  Length 425;
Best Local Similarity   29.1%;  Pred. No. 3.7e-22;
Matches   96;  Conservative   58;  Mismatches 160;  Indels  21;  Gaps   7;

Qy    25 SQTVKIFLRSLFYTALLVFSALGNIIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP 84
Db    98 SSWLTLFVPSV-YTGVFVVSLPLNIMAIVVFILKMKVKVKKPAVVYMLHLATADVLFVSVLP 156
```

```
Qy    85 GRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTA 144
Db   157 FKISYYFSGSGDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTL 216

Qy   145 GRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLM-----VLV 198
Db   217 GRASFTCLAIWALAIAGVVPLVLKEQTIQVPG---LNTTCHDVLNETLLEGYYAYFS 272

Qy   199 AF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS 257
Db   273 AFSAVFFFVPLIISTVCVSIIRCLSSSA-----VANRSKKSRALFLSAAVFCIFIICFG 327

Qy   258 PYHLNIKQFMARGMLHLPSCAERRAPLLSLQATVALMNMNCGITPIIYFFASTHYRKWLL 317
Db   328 P--TNVLLIAHYSFLSHTSTTEAAYFAYLL--CVCVSSISSCIDPLIYYASSECQRVYY 383

Qy   318 GILKLKGSSSSSSSSSTPGKASSETPSITQARGSMF 354
Db   384 SILCCKESSDPSSYNSSGQLMASKMDTCSSNLNNSIY 420


RESULT 4
A43448
thrombin receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A43448
R;Zhong, C.; Hayzer, D.J.; Corson, M.A.; Runge, M.S.
  J. Biol. Chem. 267, 16975-16979, 1992
A;Title: Molecular cloning of the rat vascular smooth muscle thrombin receptor. Evidence
A;Reference number: A43448; MUID:92381002; PMID:1324917
A;Accession: A43448
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-432 <ZHO>
A;Cross-references: GB:M81642; NID:g207465; PIDN:AAA42274.1; PID:g207466
A;Experimental source: RASM aortic smooth muscle cells
A;Note: sequence extracted from NCBI backbone (NCBIN:111973, NCBIP:111974)
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match             18.1%;  Score 339;  DB 2;  Length 432;
Best Local Similarity   29.0%;  Pred. No. 4.1e-22;
Matches   98;  Conservative   61;  Mismatches  153;  Indels  26;  Gaps   9;

Qy     6 PYFTPTKIKTKIKDLN--VESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKIN 63
Db    87 PPMPDPPFIS---EDASGYLTSPWLTLFIPSV-YTFVFIVSLPLNILAIAVFVFRMKVKK 142

Qy    64 CTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMAC 123
Db   143 PAVVYMLHLAMADVLFVSVLPFFKISYYFSGTDWQFGSGMCRFATAACYCNMYASIMLMTV 202

Qy   124 VSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEY 183
Db   203 ISIDRFLAVVYPIQSLSWRTLGRANFTCVVIWWAIMGVVPLLLKEQTTQVPG----LNI 258

Qy   184 SSMESVLGLPLM------VLVAF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKG 236
Db   259 TTCHDVLNETLLHGFYSYYFSAFSAIFFLVPLIISTVCVSIIRCLSSSA-----VANRS 313

Qy   237 RHWRGCLLTLMLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMM 296
Db   314 KKSRALFLSAAVFCIFIVCFGPTNVLL---IVHYLLLSDSPGTETAYFAYLLCVCVTSVA 370

Qy   297 NCGITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSS 334
Db   371 SC-IDPLIYYASSECQKHYSILCCRESSDSNSCNST 407


RESULT 5
I51667
thrombin receptor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
```

C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51667
R;Gerszten, R.E.; Chen, J.; Ishii, M.; Ishii, K.; Nanevicz, T.; Turck, C.W.; Vu, T.H.; C
Nature 368, 648-651, 1994
A;Title: Thrombin receptor's specificity for agonist peptide is determined by its extrac
A;Reference number: I51667; MUID:94195429; PMID:8145852
A;Accession: I51667
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-420 <GER>
A;Cross-references: EMBL:U09632; NID:g495197; PIDN:AAA18498.1; PID:g495198

Query Match   17.7%;   Score 332.5; DB 2;   Length 420;
Best Local Similarity   26.6%;   Pred. No. 1.5e-21;
Matches   92;   Conservative   74;   Mismatches 155;   Indels 25;   Gaps   8;

Qy   16 TKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVS   75
                |||::|:||: ::||::|||:|| :||: ||| : |||| :::|:|: ||:::|::|| |
Db   88 TKEABQYLSSQWLTKFVPSL-YTVFIVGLPLNLLAIIIPLFKMKVRKPAVVYMLNLAIA 146

Qy   76 DLLFTVALPGRVCVCVLGSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCA   135
                 :::|| |: |:|:|||| ::||: | | ||||: |||| ||| | ||| ||||
Db   147 DVFFVSVLPFKIAYHLSGNDWLFGPGMCRIVTA IFYCNMYCSVLLIASISVDRFLAVVYP 206

Qy   136 HWGPPLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLM   195
                : ||| | | || | :| | :| || | |:   :|::|| ||
Db   207 MHISLGWRTMSRAYMACSFIWLISIASTIPLLVTEQTQ----KIPRLDITTCHDVLDLKDL 262

Qy   196 --VLVAFAIGFC------GPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLM   248
                  :||:|:||| || :||| ||||| || :: | | |: |:| |
Db   263 KDFYIYFSSFCLLFFVPFIITTCYIGIIRSLSSSIENSC----KKTRALFLAVVV 317

Qy   249 LVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFPA   308
                :: : || ::   :| |||| | ||||: ||: || :|:||||||
Db   318 LCVFIICFGFNV--LFLIH--YLQEANEFLYPAYILSACVG--SVSCCLDPLIYYYA   369

Qy   309 STHYRKWLLGILKLKGSSSSSSSSSSSSTPGKASSETPSITQARGSMF 354
                 :| ::| :| ||:
Db   370 SSQCQRYLYSLLCCRKVSEPGSSTGQLMSTAMKDNDNCSTNAKSSIY 415

RESULT 6
JC4800
P2Y6 receptor - human
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C;Accession: JC4800; G02514
R;Communi, D.J.; Parmentier, M.; Boeynaems, J.M.
Biochem. Biophys. Res. Commun. 222, 303-308, 1996
A;Title: Cloning, functional expression and tissue distribution of the human P2Y6 recept
A;Reference number: JC4800; MUID:96222498; PMID:8670200
A;Accession: JC4800
A;Molecule type: mRNA
A;Residues: 1-328 <COM>
A;Cross-references: EMBL:X97058
A;Experimental source: placenta
R;Hammet, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01373
A;Accession: G02514
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 'M',4-328 <HAM>
A;Cross-references: EMBL:U52464; NID:g1407632; PIDN:AAB03572.1; PID:g1407633
C;Genetics:
A;Gene: P2Y6
C;Superfamily: ATP receptor P2u
C;Keywords: glycoprotein; placenta; receptor; transmembrane protein
F;26-52/Domain: transmembrane #status predicted <TMM1>
F;63-86/Domain: transmembrane #status predicted <TMM2>
F;104-122/Domain: transmembrane #status predicted <TMM3>
F;143-167/Domain: transmembrane #status predicted <TMM4>
F;193-216/Domain: transmembrane #status predicted <TMM5>

F;241-264/Domain: transmembrane #status predicted <TMM6>
F;283-305/Domain: transmembrane #status predicted <TMM7>
F;5,173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   17.5%;   Score 329; DB 2;   Length 328;
Best Local Similarity   28.0%;   Pred. No. 2.4e-21;
Matches   89;   Conservative   54;   Mismatches 141;   Indels 34;   Gaps   11;

Qy   10 PPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYL   69
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14 PPTTC-------VYRENFKQLLPPVVSAVLAAGLPNICVITIQICTSRRALTRTAVYT   65

Qy   70 VHLAVSDLLFTVALPGRVVCYVLGSSWPGKGLCRLTAPVLYTDTYGGVYLMACVSVDHY 129
                ::|||:|| ||||: | || || ||| : || | || :|| |:| | ||:| ||
Db   66 LNLALADLLYACSLPLLIYNYAQGDHWPFGDFACRLVRFLFYANLHGSILFLTCISFQRY 125

Qy   130 PAVVCAHWGPRLRTAGR--ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYS-- 184
                 |: |:| | | ||::| :: | |||| : |: | :|| | : | ||
Db   126 LG-ICHPLAPWHKRGGRRAAWLVCVAVWLAVTTQCLPTAIFAATGIQRNRTVCYDLSPPA 184

Qy   185 --SMESVLGLPLMVLVAFAIGFCGPVGIILSCVMKITWKLCSTARED----PVT-SRKGR 237
                   |||:||| |: | | | ||| |:|| ||::||||||||| |: |||
Db   185 LATHYMPYGMALTV-----IGFLLPFAALLACYCLLACRLC---RQDGPAEPVAQERRGK 236

Qy   238 HWRGCLLTLLMLVAVVVCFSPYHLNIKQFMA-RGMLHLPSCAERRAFLLSLQATVALMNM 296
                | || | :::|:: | : | | | : :|||: | | : || ::
Db   237 AAR--MAVVVAAAPAISFLPFHITKTAYLAVRSTPGVP-CTVLEAFAAAYKGTRPFFASA 292

Qy   297 NCGITPIIYFPASTHYRK 314
                ||| ||:||| || ||:
Db   293 NSVLDPLLFYFTQKKFRR 310

RESULT 7
JC5067
G protein-coupled receptor CKR-L1 - human
N;Alternate names: chemokine receptor-like protein TER1; GPR-CY6
C;Species: Homo sapiens (man)
C;Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 21-Jul-2000
C;Accession: JC5067; G02776; G02387
R;Zaballos, A.; Varona, R.; Gutierrez, J.; Lind, P.; Marquez, G.
Biochem. Biophys. Res. Commun. 227, 846-853, 1996
A;Title: Molecular cloning and RNA expression of two new human chemokine receptor-like c
A;Reference number: JC5067; MUID:97040707; PMID:8886020
A;Accession: JC5067
A;Molecule type: DNA
A;Residues: 1-355 <ZAB>
A;Cross-references: EMBL:Z79783; NID:g1668735; PIDN:CAB02142.1; PID:g1668736
R;Napolitano, M.; Zingoni, A.; Bernardini, G.; Spinetti, G.; Rocchi, M.; Santoni, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: H01714
A;Accession: G02776
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-355 <BON>
A;Cross-references: EMBL:U45983; NID:g1245056; PID:g1245057
A;Accession: G02387
A;Molecule type: DNA
A;Residues: 1-355 <NAP>
A;Cross-references: EMBL:U62556; NID:g1468978; PID:g1468979
R;Bonner, T.I.
submitted to the EMBL Data Library, January 1996
A;Reference number: H01154
A;Accession: G02387
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-355 <BON>
A;Cross-references: EMBL:U45983; NID:g1245056; PID:g1245057
C;Comment: This protein belongs to the family of beta chemokine receptors.
C;Genetics:
A;Gene: GDB:CMKBR8; CMKBRL2; TER1; CKR-L1
A;Cross-references: GDB:6053733; OMIM:601834
A;Map position: 3p21-3p21
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; transmembrane protein
F;36-63/Domain: transmembrane #status predicted <TM1>
F;73-94/Domain: transmembrane #status predicted <TM2>

```
F;108-129/Domain: transmembrane #status predicted <TM3>
F;147-171/Domain: transmembrane #status predicted <TM4>
F;200-222/Domain: transmembrane #status predicted <TM5>
F;239-260/Domain: transmembrane #status predicted <TM6>
F;281-304/Domain: transmembrane #status predicted <TM7>

Query Match             17.3%;  Score 324;  DB 2;  Length 355;
Best Local Similarity   26.4%;  Pred. No. 7.1e-21;
  Matches  89;  Conservative  65;  Mismatches 145;  Indels  38;  Gaps  8;

Qy    26 QTVKIFLRSLFYTALLVFSALGN--ILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
            |      |     ||| || |    :  :|: :|:|:| |||:||:|||||
Db    31 QTNGKLLLAVFYCLLFVFSLLGNSLVIVLVVCKKLRSI--TDVYILNLALSDLLFVFSF 88

Qy    84 PGRVVCVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT 143
            ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    89 PFQT--YYLLDQWVFGTVMCKVVSGFYYIGFYSSMFFITLMSVDRYLAVHAVYALKVRT 146

Qy   144 AGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGPLMVLVAFAIG 203
            :|:|:| :|: : ||||   : : :|:
Db   147 IRMGTTLCLAVWLTAIMATIPLLVFYQVASEDGVLQCYSFYNQQT---LKWKIFTNFKMN 203

Qy   204 FCG-*-PVGILLSCYMKITWKL--CSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSP 258
           |||     ||||||||||||    ::|||||:|:|::
Db   204 ILGLLIPTIFMFCYIKILHQLKRCQNHNKT---------KAIRLVLIVVIASLLFWVP 253

Qy   259 YHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKWL-- 316
               ||:|||  ||  :  ||| : |:||  | ::
Db   254 FNVVLFLTSLHSMHILDGCSISQQLYTATHVIEIISFTHCCVNPVIYAFVGEKFKKHLSE 313

Qy   317 ---------LGILLKGSSSSSSSSSSSSSTPGKASS 341
                     |||||||||
Db   314 IFQXSCSQIFNYLGRQMPRESCEKSSSCQQHSSRSSS 350
```

RESULT 8
A57160
chemokine (C-C) receptor 4 - human
N;Alternate names: C-C CKR-4
C;Species: Homo sapiens (man)
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C;Accession: A57160
R;Power, C.A.; Meyer, A.; Nemeth, K.; Bacon, K.B.; Hoogewerf, A.J.; Proudfoot, A.E.I.; W
J. Biol. Chem. 270, 19495-19500, 1995
A;Title: Molecular cloning and functional expression of a novel CC chemokine receptor cD
A;Reference number: A57160; MUID:95370289; PMID:7642634
A;Accession: A57160
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-360 <POW>
A;Cross-references: GB:X85740; NID:g1370103; PIDN:CAA59743.1; PID:g971452
A;Note: source clone K5-5
C;Genetics:
A;Gene: GDB:CMKBR4
A;Cross-references: GDB:677463
A;Map position: 3p21-3p21
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane prot
F;40-65/Domain: transmembrane #status predicted <TM1>
F;76-97/Domain: transmembrane #status predicted <TM2>
F;112-133/Domain: transmembrane #status predicted <TM3>
F;151-175/Domain: transmembrane #status predicted <TM4>
F;208-226/Domain: transmembrane #status predicted <TM5>
F;243-264/Domain: transmembrane #status predicted <TM6>
F;283-308/Domain: transmembrane #status predicted <TM7>
F;29-276,110-187/Disulfide bonds: #status predicted
F;72,75/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicte
F;145/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted
F;183,194/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;321/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

Query Match   17.2%; Score 322.5; DB 2; Length 360;
Best Local Similarity   28.3%; Pred. No. 9.7e-21;

```
Matches  93;  Conservative  58;  Mismatches 131;  Indels  47;  Gaps  9;

Qy    29 KIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP--GR 86
            :|:||    |   :||  ||  |:|| :  :|||||||||:|||:|||||:|||||
Db    39 ELFLPPL-YSLVFVFGLLGNSVVLVLFKYKRLRSMTDVYLLNLAISDLLFVSLPFWG-- 96

Qy    87 VVCVVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGR 146
            |  |||||||||||| | |  |||||||||||:|:    :| |
Db    97 --YYAADQWVFGLGLCKMISWMVIVGFYSGIFFVMLMSIDRYLAIVHAVFSLRARTLIY 153

Qy   147 ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCG 206
           | |  |  |    | ||  |||:||| :  :|:
Db   154 GVITSLATWSVAVFASLPGFLFSTCYTERNHTYCKTKISLNSTTWKVLSSLEINILGLVI 213

Qy   207 PVGIILSCYMKI--TWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCF----SPYH 260
            ||| |  |::|      :||||||||:|:|::
Db   214 PLGIMLFCYSMIIRTLQHCKNEKKNKAVK------------------MIFAVVVLFLGFWTPY- 258

Qy   261 LNIKQFMAR--GMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKWLL- 317
             |:|:|    : :  |  |  :|:||||  ||  ::
Db   259 -NIVLFLETLVELEVLQDCTFERYLDYAIQATETLAFVHCCLNPIIYFFFLGEKFRKYILQ 317

Qy   318 ---------------GILKLKGSSSSSSSS 331
                           ||||||||||||||||
Db   318 LFKTCRGLFVLCQYCGLLQIYSADTPSSS 346
```

RESULT 9
JC5498
G protein-coupled receptor DEZ - mouse
C;Species: Mus musculus (house mouse)
C;Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 19-May-2000
C;Accession: JC5498
R;Methner, A.; Hermey, G.; Schinke, B.; Hermans-Borgmeyer, I.
Biochem. Biophys. Res. Commun. 233, 336-342, 1997
A;Title: A novel G protein-coupled receptor with homology to neuropeptide and chemoattr
A;Reference number: JC5498; MUID:97289630; PMID:9144535
A;Contents: Brain
A;Accession: JC5498
A;Molecule type: mRNA
A;Residues: 1-371 <MET>
A;Cross-references: GB:U79525; NID:g1732346; PIDN:AAB53789.1; PID:g1732347
C;Comment: This protein is involved in the bone metabolism.
C;Superfamily: vertebrate rhodopsin
F;110-187/Disulfide bonds: #status predicted

Query Match   16.8%; Score 315.5; DB 2; Length 371;
Best Local Similarity   25.4%; Pred. No. 4.1e-20;

```
Matches  89;  Conservative  67;  Mismatches 154;  Indels  41;  Gaps  9;

Qy    24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
               :  |:|  || | || | || ||| |  |:     :   |||:|||:|||||||||
Db    33 EAKVAPVFL-VVIYSLVCFLGLLGNGLVIVIATFKMKKTVNT-VWFVNLAVADFLFNIPL 90

Qy    84 PGRVVCVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT 143
            |||| |||  || |||| | |  |||||||||||:| :||    : |:
Db    91 PMHITYAAMDYHWVFGKAMCKISNFLLSHMYMTSVFLLTVISFDRCISVLLPVWSQNHRS 150

Qy   144 AGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLP---------- 193
            |||| | ||  :  ||     |||:||| :
Db   151 IRLAYMTCSAVWVLAFFLSSPSLVFRDTANIHGKITCFNNFSLAAPESSPHPAHSQVVST 210

Qy   194 ------LMVLVAFAIGFCGPVGIILSCYMKITWKLTWKLCSTAREDPVTSRKGRHWRGCLLTLL 247
                  |||:|||   ||  ||||||||||
Db   211 GYSRHVAVTVTRFLCGFLIPVFIITACYLTIVFKL---QRNRLAKNKKPFKIIT 262

Qy   248 MLVAVVVCFSPYH-LNIKQFMARGMLH--LPSCAERRAFLLSLQATVALMNNCGITPI 303
            :|: |||| :||  |||||||
Db   263 IIITFFPLCWCPYHTLYLLB-----LHHTAVPS----SVFSLGLPLATAVAIANSCMNPI 312

Qy   304 IYFFASTHYRKWLLLGIL-KLKGSSSSSSSSSSSSSTPGKASSETPSITQARGSM 353
            ||||||||||||||:|
Db   313 LVVFMGHDFDRKFKVALPSRLANALSEDTGPSSYPSHRSFTKMSSLNEKASV 363
```

RESULT 10
I55450
G protein-coupled P2 receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Jul-1996 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I55450
R;Chang, K.; Hanaoka, K.; Kumada, M.; Takuwa, Y.
J. Biol. Chem. 270, 26152-26158, 1995
A;Title: Molecular cloning and functional analysis of a novel P2 nucleotide receptor.
A;Reference number: I55450; MUID:96064682; PMID:7592819
A;Accession: I55450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA <RES>
A;Residues: 1-328 <RES>
A;Cross-references: GB:D63665; NID:g1066007; PIDN:BAA09816.1; PID:g1066008
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor

Query Match   16.4%; Score 308; DB 2; Length 328;
Best Local Similarity 25.6%; Pred. No. 1.7e-19;
Matches 80; Conservative 58; Mismatches 152; Indels 22; Gaps 8;

Qy   10 PPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIALCLTCQKSRKINCTGIYL 69
Db   14 PPTTC-------VYREDFKRLLLPPVYSVVLLVVGLPLNVCVIAQICASRRTLTRSAVYT 65

Qy   70 VHLAVSDLLFTVALPGRVVCVYLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSSVDHY 129
Db   66 LNLALADLLYACSLPLLIYNYARGDHWPFGDLACRLVRFLFYANLHGSILFLTCISFQRY 125

Qy 130 PAVVCAHWGPRLRTAGR--ARLVCVAIWTLVLLQTMPLLMPMTKPLVGKLACMEYS-SM 186
Db 126 LG-ICHPLAPWHKRGGRRAAWTVCGVVMLVVTAQCLPTAVPAATGIQRNRTVCYDLSPPI 184

Qy 187 ESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTARED----PVTSRKGRHWRGC 242
Db 185 LSTRYLPYGMALT-VIGFLLPFTALLACYCRMARRLC---RQDGPAGPVAQE--RRSKAA 238

Qy 243 LLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITP 302
Db 239 RMAVVVAAVFVISFLPFHITKTAYLAVRSTPGVSCPVLETFAAAYKGTRPPASANSVLDP 298

Qy 303 IIYFFASTHYRK 314
Db 299 ILFYFTQQKFRR 310

RESULT 11
A47556
ATP receptor P2u - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C;Accession: A47556
R;Lustig, K.D.; Shiau, A.K.; Brake, A.J.; Julius, D.
Proc. Natl. Acad. Sci. U.S.A. 90, 5113-5117, 1993
A;Title: Expression cloning of an ATP receptor from mouse neuroblastoma cells.
A;Reference number: A47556; MUID:93281707; PMID:7685114
A;Accession: A47556
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-373 <LUS>
A;Cross-references: GB:L14751; NID:g309457; PIDN:AAA39871.1; PID:g309458
C;Superfamily: ATP receptor P2u
C;Keywords: transmembrane protein

Query Match   16.4%; Score 308; DB 2; Length 373;
Best Local Similarity 27.2%; Pred. No. 1.9e-19;
Matches 96; Conservative 53; Mismatches 164; Indels 40; Gaps 8;

Qy   25 SQTVKIFLRSLFYTALLVFSALGNIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP 84
Db   28 NEDFKVLLPVSYGVVCVLGLCLNVVALYIFLCRLKTWNASTTYMFHLAVSDLYAASLP 87

Qy   85 GRVVCVYLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVV----CAHWGPR 140
Db   88 LLVYYYARGDHWPFSTVLCKLVRFLFYTNLYCSILFLTCISVHRCLGVLRPLHSLRWG-- 145

Qy 141 LRTAGRARLVCVAIWTLVLLQTMPLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAF 200
Db 146 --RARYARRVAAVVWTVLVLACQAPVLYFVTTSVRGTRITCHDTSARELFSHFVAYSSVML 203

Qy 201 AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAV-VVCFSPY 259
Db 204 GLLFAVPFSVIIVCYVLMARRLLKPAY--GTTGGLPRAKRKSVRTIALVLAVFALCFLPF 261

Qy 260 HLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWL- 316
Db 262 HVTRTLYYSFRSLDI-SCHTLNAINMAYKITRPLASANSCLDPVLYFLAGQRLVFARDA 320

Qy 317 -----------LGILK-------LKGSSSSSSSSTPGKASSETPSI 346
Db 321 KPPTEPTSPQARRKLGLHRPNRTVRKDLSVSSDDSRRTESTP--AGSETKDI 371

RESULT 12
A39714
G protein-coupled receptor RDC1 - human
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 24-Nov-1999
C;Accession: A39714
R;Sreedharan, S.P.; Robichon, A.; Peterson, K.E.; Goetzl, E.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 4986-4990, 1991
A;Title: Cloning and expression of the human vasoactive intestinal peptide receptor.
A;Reference number: A39714; MUID:91271317; PMID:1675791
A;Accession: A39714
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-362 <SRE>
A;Cross-references: GB:M64749; NID:g292418; PID:g292419
C;Comment: The authors have revised the name and functional description of this protein
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   16.4%; Score 307.5; DB 2; Length 362;
Best Local Similarity 24.2%; Pred. No. 2e-19;
Matches 80; Conservative 74; Mismatches 148; Indels 29; Gaps 6;

Qy   22 NVESQTVKIFLRSLFYTALLVFSALGNIALCLTCQKSRKINCTGIYLVHLAVSDLLFTV 81
Db   36 NMPNKSVLLYTLSFIYIFIFVIGMIANSVVWVNIQAKTGYDTHCYILNLAIADLWVL 95

Qy   82 ALPGRVVCVYLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRL 141
Db   96 TIPVWVVSLVQHNQWPMGELTCKVTHLIFSINLFSGIFFLTCMSVDRYLSITYFNTPSS 155

Qy 142 RTAGRARLVCVAIWTLVLLQTMP-LLLMPMTKPLVGKLACM-----EYSSMESVLGLPLM 195
Db 156 RKKMVRRVCILVWLLAFCVSLPDTIYLKTVTSASNNETYCRSFYPEHSIKEWLIG--M 212

Qy 196 VLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVC 255
Db 213 ELVSVVLGFAVPFSIIAVFYFLLARAISASSDQEKHSSRK-----IIFSYVVFLVC 264

Qy 256 FSPYHLNIKQFMARGMLHLP-SCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRK 314
Db 265 WLPYHVAVLLDIFSILHYIPFTCRLEHALFTALHVTQCLSLVHCCVNPVLYSFINRNYRY 324

Qy 315 WLL----------GILKLKGSSSSSSSSS 334
Db 325 ELMKAFIFKYSAKTGLTKLIDASRVSETEYS 355

RESULT 13
A45177

interleukin-8 receptor type B - human
N;Alternate names:
C;Species: Homo sapiens (man)
C;Date: 07-Oct-1994 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: I37898, I38712; A53611; A39446
R;Ahuja, S.K.; Shetty, A.; Tiffany, H.L.; Murphy, P.M.
J. Biol. Chem. 269, 26381-26389, 1994
A;Title: Comparison of the genomic organization and promoter function for human interleul
Cell 72, 415-425, 1993
A;Reference number: I37898; MUID:95014476; PMID:7929358
A;Accession: I37898
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <RES>
A;Cross-references: EMBL:U11869; NID:g511801; PIDN:AAB60656.1; PID:g511803
A;Accession: I38712
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-15 <RE2>
A;Cross-references: EMBL:U11872; NID:g511808; PIDN:AAA64380.1; PID:g511809; EMBL:U11873;
11876; NID:g511816; PID:g511817; EMBL:U11877; NID:g511818; PID:g511819; EMBL:U11878; NID
R;Sprenger, H.; Lloyd, A.R.; Lautens, L.L.; Bonner, T.I.; Kelvin, D.J.
J. Biol. Chem. 269, 11065-11072, 1994
A;Title: Structure, genomic organization, and expression of the human interleukin-8 rece
A;Reference number: A53611; MUID:94209273; PMID:7512557
A;Accession: A53611
A;Status: preliminary
A;Molecule type: DNA
A;Genetics:
A;Residues: 6-360 <SPR>
A;Cross-references: GB:M99412; GB:L19593
R;Murphy, P.M.;Tiffany, H.L.
Science 253, 1280-1283, 1991
A;Title: Cloning of complementary DNA encoding a functional human interleukin-8 receptor
A;Reference number: A39446; MUID:91368200; PMID:1891716
A;Accession: A39446
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 6-360 <MUR>
A;Cross-references: GB:M73969
C;Comment: This receptor, unlike IL8RA, binds several peptides besides interleukin-8, in
C;Genetics:
A;Gene: GDB:IL8RB; IL8RA
A;Cross-references: GDB:127868; OMIM:146928
A;Map position: 2q35-2q35
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match   16.3%;   Score 306.5; DB 2; Length 360;
Best Local Similarity 28.0%; Pred. No. 2.5e-19;
Matches 94; Conservative 61; Mismatches 144; Indels 37; Gaps 9;

```
Qy   24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL   83
Db   42 ESLEINKYFVVIIYALVFLLSLLGNSLVMLVILYSRVGRSVTDVVLLNLALADLLFALTL  101
Qy   84 PGRVVCVLGSSWPFGKGLCRLTAFVLYTDTYGGVVLMACVSVDHYPAVVCAHWGPRLRT   143
Db  102 PIWAASKVNG--WIFGTFLCKVVSLLACISVDRYLAIVHA---TRTLT            156
Qy  144 AGR--ARLVCVAIWTLVLLQTMPLLMPMT-----KPLVGKLACMEYSSMESVLGLPLM   195
Db  157 QKRYLVKFICLSIWGLSLLLALPVLLFRRTVSSNVSP------ACYEDMGNNTANWRMLL  211
Qy  196 VLVAFAIGFCGPVGIILSCY---MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAV  252
Db  212 RILPQSFGFIVPLLIMLFCYGFTLRRTLFK-------AHMGQKIRAMRVIFAVVLIF    260
Qy  253 VVCFSPYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTH  311
Db  261 LLCWLPYNLVLLADTLMRTQVIQETCERRNHIDRALDATEILGILHSCLNPLIYAFIGQK  320
Qy  312 YRKWLLGILKLKGSSSSSSS----SSSSTPGKASSET                        343
Db  321 FRHGLLKILAIHGLISKDSLPKDSRPSFVGSSSGHT                         356
```

---

chemokine (C-C) receptor 1 - human
N;Alternate names: C-C CKR-1; macrophage inflammatory protein-1-alpha receptor
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
C;Accession: A45177; I55671
R;Neote, K.; DiGregorio, D.; Mak, J.Y.; Horuk, R.; Schall, T.J.
Cell 72, 415-425, 1993
A;Title: Molecular cloning, functional expression, and signaling characteristics of a C-
A;Reference number: A45177; MUID:93161416; PMID:7693328
A;Accession: A45177
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-355 <NEO>
A;Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
A;Experimental source: HL60 cells
A;Note: sequence extracted from NCBI backbone (NCBIP:124876)
R;Gao, J.
J. Exp. Med. 177, 1421-1427, 1993
A;Title: Structure and functional expression of the human macrophage inflammatory 1 alph
A;Reference number: I55671; MUID:93240122; PMID:7683036
A;Accession: I55671
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-355 <RES>
A;Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
C;Genetics:
A;Gene: GDB:CMKBR1; CMKR-1
A;Cross-references: GDB:138446; OMIM:601159
A;Map position: 3p21-3p21
C;Superfamily: vertebrate rhodopsin
C;Keywords: disulfide bond; G protein-coupled receptor; glycoprotein; phosphoprotein; tr
F;36-60/Domain: transmembrane #status predicted <TM1>
F;71-91/Domain: transmembrane #status predicted <TM2>
F;108-129/Domain: transmembrane #status predicted <TM3>
F;147-171/Domain: transmembrane #status predicted <TM4>
F;205-223/Domain: transmembrane #status predicted <TM5>
F;240-264/Domain: transmembrane #status predicted <TM6>
F;288-305/Domain: transmembrane #status predicted <TM7>
F;5/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;24-273,106-183/Disulfide bonds: #status predicted
F;345/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted

Query Match   16.4%;   Score 307; DB 2; Length 355;
Best Local Similarity 27.3%; Pred. No. 2.2e-19;
Matches 86; Conservative 58; Mismatches 143; Indels 28; Gaps 5;

```
Qy   37 YTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVLGSSW   96
Db   41 YSLVFVIGLVGNILVVLVLVQYRLKNMTSIYLLNLAISDLLFLFTLP-FWIDYKLKDDW   99
Qy   97 PFGKGLCRLTAFVLYTDTYGGVVLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWT  156
Db  100 VFGDAMCKILSGFYYTGLYSEIFFIILLTIDRYLAIVHAVFALRARTVTFGVITSIIIWA  159
Qy  157 LVLLQTMPLLMPMTKPLVGKLACMEYSSMESVLGLPL-MVLVAFAIGFCGPVGIILSCY   215
Db  160 LAILASMPGLYFSKTQWEFTHHTCSLHFPHESLREWKLFQAALKLNLFGLVLPLLVMIICY  219
Qy  216 MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVVCFSPYHLNIKQFMARGMLHLP  275
Db  220 TGIIKILLRRPNE-----KKSKAVR---LIFVIMIIFFLFWTPYNLTILISVFQDFLFTH   271
Qy  276 SCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWL------L              317
Db  272 ECEQSRHLDLAVQVTEVIAYTHCCVNPVIYAFVGERFRKYLRQLFHRRVAVHLVKWLPFL  331
Qy  318 GILKLKGSSSSSSSS                                               332
Db  332 SVDRLERVSSTSPST                                               346
```

RESULT 14
A53611

RESULT 15

I38435

angiotensin receptor homolog APJ - human

C;Species: Homo sapiens (man)

C;Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 21-Jul-2000

C;Accession: I38435

R;O'Dowd, B.F.; Heiber, M.; Chan, A.; Heng, H.H.; Kennedy, J.L.; Tsui, L.; Kennedy, J.L.; Shi, X.; Petr

Gene 136, 355-360, 1993

A;Title: A human gene that shows identity with the gene encoding the angiotensin recepto

A;Reference number: I38435; MUID:94124031; PMID:8294032

A;Accession: I38435

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-380 <RES>

A;Cross-references: EMBL:U03642; NID:g425351; PIDN:AAA18954.1; PID:g425352

C;Genetics:

A;Gene: APJ

A;Map position: 11q12

A;Introns: #status absent

C;Superfamily: vertebrate rhodopsin

```
Query Match         16.1%;  Score 302;  DB 2;  Length 380;
Best Local Similarity  26.7%;  Pred. No. 6.5e-19;
Matches   88;  Conservative  67;  Mismatches 152;  Indels  22;  Gaps  9;

Qy    37  YTALLVFSALGNILALCLTCCKSR-KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS  95
               :  ||   ||  || :| :| ||||  :|: :||:|||
Db    35  YMLVFLLGTTGNGLVLWTVFRSSREKRRSADIFIASLAVADLTFVVTLPLWATYTYRDYD  94

Qy    96  WPFGKGLCRLTAFVLYTDTYGGVLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW  155
              |||  ||:  |:::::::  | : :|| |  ||:|
Db    95  WPFGTFFCKLSSYLIFVNMYASVFCLTGLSFDRYLAIVRPVANARLRLRVSGAVATAVLW  154

Qy   156  TLVLLQTMPLLLMPMTKPL--VGKLAC-MEYSSMESV------LGLPLMVLVAPAIGFC  205
              |||:||    ::| ||: |:|:|  |||| :||  :||
Db   155  VLAALLAMPVMVLRTTGDLENTTKVQCYMDYSMVATVSSEWAWEVGLGVS---STTVGFV  211

Qy   206  GPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAV--VVCFSPYHLNI  263
              :|:|| :||  :| |: ::::  |    ||:  |:|  :: ||||
Db   212  VPFTIMLTCYFFIAQTIAGHFRKERI--EGLRKRRLLSIIVLVVTFALCWMPYHLVK  268

Qy   264  KQFMARGMLHLPSCAERRAFLLSL-QATVALMNMNCGITPIIYFFASTHYRKWLLGILKL  322
              :  ||| :|||:  ::  | ||  :|: :|     |: ::|  |:
Db   269  TLYMLGSLLHWP--CDFDLFLMNIFPYCTCISYVNSCLNPFLYAFFDPRFRQACTSMLCC  326

Qy   323  KGSSSSSSSSTPGKASSETPSITQARG  351
              :::|||:  :|:| |:|
Db   327  GQSRCAGTSHSSSGEKSASYSSGHSQGPG  355
```

Search completed: February 19, 2004, 19:34:56

Job time : 22 secs

```
                      GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 19, 2004, 19:18:00 ; Search time 17 Seconds
                              (without alignments)
                              995.860 Million cell updates/sec

Title:        US-10-041-615-34
Perfect score: 1876
Sequence:   1 MIKLGPYFPPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:   127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:   127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	632.5	33.7	361	1	EBI2_HUMAN	P32249 homo sapien
2	391.5	20.9	367	1	GP17_HUMAN	Q13304 homo sapien
3	367	19.6	428	1	PAR1_CRILO	Q00991 cricetulus
4	358.5	19.1	425	1	PAR1_PAPHA	P56488 papio hamad
5	357		430	1	PAR1_MOUSE	P30558 mus musculu
6	352.5	18.8	425	1	PAR1_HUMAN	P25116 homo sapien
7	341	18.2	308	1	V206_FOWPV	Q9J529 fowlpox vir
8	339	18.1	432	1	PAR1_RAT	P26824 rattus norv
9	332.5	17.7	420	1	PAR1_XENLA	P47749 xenopus lae
10	332	17.7	537	1	P2Y8_XENLA	P79928 xenopus lae
11	329	17.5	328	1	P2Y6_HUMAN	Q15077 homo sapien
12	326	17.4	353	1	CKR8_MOUSE	P56484 mus musculu
13	324	17.3	355	1	CKR8_HUMAN	P51685 homo sapien
14	323.5	17.2	340	1	CLT1_PIG	Q95n02 sus scrofa
15	322.5	17.2	360	1	CKR4_HUMAN	P51679 homo sapien
16	322	17.2	337	1	CLT1_HUMAN	Q9y271 homo sapien
17	318	17.0	339	1	CLT1_RAT	Q924t8 rattus norv
18	317	16.9	355	1	CKR1_MACMU	P56482 macaca mula
19	316.5	16.9	333	1	FK79_HUMAN	Q9bxc1 homo sapien
20	316	16.8	352	1	CLT1_MOUSE	Q99ja4 mus musculu
21	316	16.8	373	1	CML1_HUMAN	Q99788 homo sapien
22	315.5	16.8	371	1	CML1_MOUSE	P97468 mus musculu
23	308	16.4	328	1	P2Y6_RAT	O63371 rattus norv
24	308	16.4	345	1	CLT2_PIG	Q95n03 sus scrofa
25	308	16.4	373	1	P2Y2_MOUSE	P35383 mus musculu
26	307.5	16.4	353	1	IL8B_PANTR	Q28807 pan troglod
27	307.5	16.4	362	1	RDC1_HUMAN	P25106 homo sapien
28	307.5	16.4	377	1	P2Y2_HUMAN	P41231 homo sapien
29	307	16.4	355	1	CKR1_HUMAN	P32246 homo sapien
30	306.5	16.3	360	1	IL8B_HUMAN	P25025 homo sapien
31	305.5	16.3	374	1	P2Y2_RAT	P41232 rattus norv
32	304.5	16.2	353	1	IL8B_MACMU	Q28519 macaca mula
33	302	16.1	367	1	CCR3_MOUSE	O88410 mus musculu
34	302	16.1	380	1	APJ_HUMAN	P35414 homo sapien
35	301	16.0	350	1	IL8A_GORGO	P55919 gorilla gor
36	300.5	16.0	371	1	CML1_RAT	O35786 rattus norv
37	299	15.9	368	1	CCR3_HUMAN	P49682 homo sapien
38	298.5	15.9	358	1	IL8B_RABIT	P35344 oryctolagus
39	298.5	15.9	359	1	IL8B_MOUSE	P35343 mus musculu
40	298	15.9	360	1	IL8B_BOVIN	Q28003 bos taurus
41	298	15.9	370	1	P2Y9_HUMAN	Q99677 homo sapien
42	297.5	15.9	308	1	P2Y5_CHICK	P32250 gallus gall
43	297.5	15.9	353	1	IL8B_GORGO	Q28422 gorilla gor
44	297	15.8	355	1	IL8A_RABIT	P21109 oryctolagus
45	297	15.8	380	1	APJ_MACMU	O97666 macaca mula

ALIGNMENTS

```
RESULT 1
EBI2_HUMAN
ID   EBI2_HUMAN      STANDARD;      PRT;    361 AA.
AC   P32249;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   EBV-induced G protein-coupled receptor 2 (EBI2).
GN   EBI2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93188173; PubMed=8383238;
RA   Birkenbach M.P., Josefsen K., Yalamanchili R.R., Lenoir G.M.,
RA   Kieff E.;
RT   "Epstein-Barr virus-induced genes: first lymphocyte-specific G
RT   protein-coupled peptide receptors.";
RL   J. Virol. 67:2209-2220(1993).
CC   -!- FUNCTION: ORPHAN RECEPTOR. PROBABLE MEDIATOR OF EBV EFFECTS ON B
CC       LYMPHOCYTES OR OF NORMAL LYMPHOCYTE FUNCTIONS.
CC   -!- TISSUE SPECIFICITY: B-LYMPHOCYTE CELL LINES.
CC   -!- INDUCTION: BY EBV.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L08177; AAA35924.1; -.
DR   PIR; B45680; B45680.
DR   Genew; HGNC:3128; EBI2.
DR   MIM; 605741; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. ..; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     31   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     32     57   1 (POTENTIAL).
FT   DOMAIN       58     77   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     78     95   2 (POTENTIAL).
FT   DOMAIN       96    105   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    106    127   3 (POTENTIAL).
FT   DOMAIN      128    149   CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM     150    168
FT   DOMAIN       169    192     4 (POTENTIAL).
FT   TRANSMEM     193    215     EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       216    241     5 (POTENTIAL).
FT   TRANSMEM     242    265     CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       266    287     6 (POTENTIAL).
FT   TRANSMEM     288    312     EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       313    361     7 (POTENTIAL).
FT   DOMAIN       126    134     CYTOPLASMIC (POTENTIAL).
FT   DISULFID     104    181     INTERACTION WITH G PROTEINS.
FT                               BY SIMILARITY.
SQ   SEQUENCE     361 AA;   41224 MW;   B5A2171F34C9C67B CRC64;

Query Match    33.7%;   Score 632.5;   DB 1;   Length 361;
Best Local Similarity   37.8%;   Pred. No. 5.3e-38;
Matches 135;   Conservative  74;   Mismatches 141;   Indels   7;   Gaps    4;

Qy    2 IKLGPVFTPPTKI-KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR   60
          |::   ||||   :
Db    3 IQMANNFTPPSATPQGNDCDLYAHHSTARIVM-PLHYSLVFIIGLVGNLLALVVIVQNRK   61

Qy   61 KINCTGIYLVHLLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL  120
         ||| :: ::  : |  |||||  :   ::: |||
Db   62 KINSTTLYSTNLVISDILFTTALPTRIAYYAMGFDWR.IGDALCRITALVFYINTYAGVNF  121

Qy  121 MACVSVDHYPAVVCCAHWGPRLRTAGRARLVCVAIWTIVLLQTMPLLLMPMTKPLVGKLAC  180
         | | || |   |     :::    :  |  ||| :::  ::|||||||||||:
Db  122 MTCLSIDRFIAVVHPLRYNKIKRIEHAKGVCIFWILVPAQTLPLLLINPMSKQEAERITC   181

Qy  181 MEYSSMESVLGLPLMVLVAPAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR   240
         |||  ::  ||| :||  | |  | |||| ::||  :  ||
Db  182 MEYPNFEETKSLPWILLGACFIGYVLPLIIILICYSQICCKLFRTAKQNPLTEKSGVNKK   241

Qy  241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGM--LHLPSCAERRAFLLSLQATVALMNMNC   298
          | : :: ||| : |  |||| :  :       : ::   :  ::: |   ::  | |||
Db  242 ALNTILLIIVVFVLCFTPYHVAIIQHMIKKLRFSNFLECSQRHSFQISLHFTVCLMNFNC   301

Qy  299 GITPIIYFFASTHYRKWLLGGILKLKGSSSSSSSSTPGKAS---SETPSITQARGS      352
          | |:|||||    :|:   :   :                     : || :::
Db  302 CMDPFIYFFACKGYKRKVMRMLKRQVSVSISSAVKSAPENSREMTETQMMIHSKSS      358
```

```
RESULT 2
GP17_HUMAN
ID   GP17_HUMAN     STANDARD;     PRT;     367 AA.
AC   Q13304; Q9UDZ6; Q9U221;
DT   01-NOV-1997 (Rel. 35, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable P2Y purinoceptor GPR17 (P2y-like receptor) (R12).
GN   GPR17.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=96145150; PubMed=8558062;
RA   Raport C.J., Schweickart V.L., Chantry D., Eddy R.L. Jr., Shows T.B.,
RA   Godiska R., Gray P.W.;
RT   "New members of the chemokine receptor gene family.";
RL   J. Leukoc. Biol. 59:18-23(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Hippocampus;
RX   MEDLINE=98181695; PubMed=9523551;
RA   Blaesius R.H., Weber R.G., Lichter P., Ogilvie A.;
RT   "A novel orphan G protein-coupled receptor primarily expressed in the
RT   brain is localized on human chromosomal band 2q21.";
RL   J. Neurochem. 70:1357-1365(1998).
CC   -!- FUNCTION: Putative receptor for purines coupled to G-proteins (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q13304-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q13304-2; Sequence=VSP_001987;
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
DR   EMBL; U33447; AAB16746.1; -.
DR   EMBL; Y12546; CAA73144.1; -.
DR   EMBL; Z94154; CAB08107.1; -.
DR   EMBL; Z94155; CAB08108.1; -.
DR   HSSP; P34996; 1DDD.
DR   Genew; HGNC:4471; GPR17.
DR   MIM; 603071; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004950; F:chemokine receptor activity; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin...  .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Glycoprotein;
KW   Alternative splicing;
FT   DOMAIN         1     64     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      65     85     1 (POTENTIAL).
FT   DOMAIN        86     92     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      93    113     2 (POTENTIAL).
FT   DOMAIN       114    133     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     134    154     3 (POTENTIAL).
FT   DOMAIN       155    175     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     176    196     4 (POTENTIAL).
FT   DOMAIN       197    223     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     224    244     5 (POTENTIAL).
FT   DOMAIN       245    260     CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     261    281     6 (POTENTIAL).
FT   DOMAIN       282    308     EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     309    329     7 (POTENTIAL).
FT   DOMAIN       330    367     CYTOPLASMIC (POTENTIAL).
FT   DISULFID     132    209     BY SIMILARITY.
FT   CARBOHYD      42     42     N-LINKED (GLCNAC...  )   (POTENTIAL).
FT   CARBOHYD     204    204     N-LINKED (GLCNAC...  )   (POTENTIAL).
FT   CARBOHYD     282    282     N-LINKED (GLCNAC...  )   (POTENTIAL).
FT   VARSPLIC       1     28     Missing (in isoform 2).
FT                               /FTId=VSP_001987.
SQ   SEQUENCE     367 AA;   40989 MW;   132FB97BEB3C60C CRC64;

Query Match    20.9%;   Score 391.5;   DB 1;   Length 367;
Best Local Similarity   30.5%;   Pred. No. 5.8e-21;
Matches  95;   Conservative  57;   Mismatches 142;   Indels  17;   Gaps    5;

Qy   36 FYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLLAVSDLLFTVALPGRVVCYVLGSS   95
         |||   : :||| :||| :|      ||   |:|: |||||| |  :||:
Db   65 FYLLDFILALVGNTLALMLFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH   124

Qy   96 WPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCCAHWGPRLRTAGRARLVCVAIW  155
         ||||  :||| |  :| : | :  |  ::: |::    ::    :|:::    |  ::  :
Db  125 WPFGBIACRLTGFLFLYLNMYASIYFJTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW  184

Qy  156 TLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVLVAFAIGFCGPVGIILSC   214
          :|  | :|||:      :       :  :: |:  :  |::|::|
Db  185 VVVAMAPLLVSPQTVQTNHTVCLQLYREKAS----HHALVSLAVAFTPFITTVTC       239

Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL   274
```

```
Db    240  YLLIIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG  294
                :|:      |   |||:||:  |||:     ::||||  |||||||     ::|      ::
Qy    275  PSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKWLLGII---KLKG--SSSS    328
Db    295  ASCATQRILALANRITSCLTSLNGALDPIMYFVAEKFRHALCNLLCGKRLKGPPPSFEG   354
                :|:                                          :|| :
Qy    329  SSSSSSTPGKA  339
Db    355  KTNESSLSAKS  365
```

RESULT 3
PAR1_CRILO
ID   PAR1_CRILO   STANDARD;   PRT;   428 AA.
AC   Q00991; Q60461;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   F2R OR PAR1.
OS   Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10030;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=91348247; PubMed=1652467;
RA   Rasmussen U.B., Vouret-Craviari V., Jallat S., Schlesinger Y.,
RA   Pages G., Pavirani A., Lecocq J.P., Pouyssegur J.,
RA   Obberghen-Schilling E.;
RT   "cDNA cloning and expression of a hamster alpha-thrombin receptor
RT   coupled to Ca2+ mobilization.";
RL   FEBS Lett. 288:123-128(1991).
RN   [2]
RP   SEQUENCE OF 42-428 FROM N.A.
RX   MEDLINE=96028007; PubMed=7488069;
RA   Hartmann T., Grace M.B., Buzard G.S., Ruoss S.J.;
RT   "Thrombin receptor polymorphism in Chinese hamster.";
RL   Biochem. Biophys. Res. Commun. 215:974-980(1995).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC   G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC   functions as a tethered ligand.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --
CC   --
DR   EMBL; X61958; CAA43957.1; -.
DR   EMBL; U34047; AAA86747.1; -.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation; Polymorphism.
FT   SIGNAL   1   21   POTENTIAL.
FT   PROPEP   22   41   REMOVED FOR RECEPTOR ACTIVATION (BY
FT   SIMILARITY).
FT   CHAIN   42   428   PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN   42   105   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   106   131   1 (POTENTIAL).
FT   DOMAIN   132   140   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   141   160   2 (POTENTIAL).
FT   DOMAIN   161   179   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   180   201   3 (POTENTIAL).
FT   DOMAIN   202   221   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   222   242   4 (POTENTIAL).
FT   DOMAIN   243   271   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   272   291   5 (POTENTIAL).
FT   DOMAIN   292   314   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   315   337   6 (POTENTIAL).
FT   DOMAIN   338   352   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   353   377   7 (POTENTIAL).
FT   DOMAIN   378   428   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   60   63   ASP/GLU-RICH (ACIDIC).
FT   DOMAIN   301   304   POLY-SER.
FT   SITE   41   42   CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID   178   257   BY SIMILARITY.
FT   CARBOHYD   65   65   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   78   78   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   253   253   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   262   262   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT   84   84   MISSING (IN AN ALLELE).
FT   CONFLICT   384   384   H -> T (IN REF. 1).
SQ   SEQUENCE   428 AA; 47602 MW; E5492AB2CD877E2F CRC64;

Query Match   19.6%; Score 367; DB 1; Length 428;
Best Local Similarity   30.1%;
Matches 102;   Conservative   63; Mismatches 158; Indels   16; Gaps   8;
Pred. No. 3.6e-19;
```
Qy      2  IKLGPVFTPPTKIKTKI-KDLN--VESQTVKIFLRSLFYTALLVFSALGNILALCLTCQK   58
               :||  |  :  |||:||   |:|      :|::|| | ::|| : | | :|
Db     75  IYLNKSHSPPAPLAPPISEDASGYLTSPWLRLFIPSV-YTFVFVVSLPLNILAIAVFVLK  133

Qy     59  SRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGV  118
               :||  :::||:|||| |:||| :  |   |  |:   : |  :|  : ||    ::|
Db    134  MKVKPAVVVMLHLAMADVLFVSVLPLKISYYFSGSDWQFGSGMCRFATAAFYCNMYASI   193

Qy    119  YLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTK--PLVG  176
               :||:  :|::  |:| |    ||  |:     |  | :|::   || ||   :|   |||
Db    194  MLMTVISIDRFLAVVYPIQSLSWRTLGRANFTCLVIWVMAIMGVVPLLLKEQTTRVPGLN  253

Qy    177  KLACMEYSSMESVLGLPLMVLVAF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRK  235
               ::|    : :   |  |:|:  | | | ||| | ::|   ::     |:
Db    254  ITTCHDVLNETLLQGFSYSFSAFSAVFFLVPLIISTICVMSIIRCLSSSS-----VANR   308

Qy    236  GRHWRGCLLITLLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMN  295
               |::    || |  : |:   |  |:    :|      |:|||| :| :|  | ::
Db    309  SKKSRALFLSAAVFCVFIVCFGP--TNVLLIMHYLLLSDSPATEKAYFAYLL--CVCVSS  364

Qy    296  MNCGITPIIYFFASTHYRKWLLGLILKLKGSSSSSSSSSS  334
               ::| | || |::   |  :||:   :::
Db    365  VSCCIDPLIYYASSECQRHLYGILCCKESSDPNSYNST   403
```

RESULT 4
PAR1_PAPHA
ID   PAR1_PAPHA   STANDARD;   PRT;   425 AA.
AC   P56488;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   F2R OR PAR1 OR BTHR12.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shoji M., Hayzer D.J., Hanson S.R.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
```

```
Db   328  P--TNILIAHYSFLSHTSTTRAAYPAYLL--CVCVSSISCCIDPLIYYASSECQRYVV 383
                :   :  ::: : :|:::::| :|:  ::|:|   :: :::|| :::::|| :||||
Qy   318  GILKLKGSSSSSSSSSSSTPGKASSETPSITQARGSMF 354
                :||::::||||||||||||||::|:||::|||
Db   384  SILCCKESSDPSSSNSSGQLMASKMDTCSSNLNNSIY 420
```

RESULT 5
PAR1_MOUSE
```
ID   PAR1_MOUSE      STANDARD;      PRT;     430 AA.
AC   P30558; P97507;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   F2R OR PAR1 OR CF2R.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Coughlin S.R.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=96379236; PubMed=8784787;
RA   Kahn M.L., Ishii K., Kuo W.L., Piper M., Connolly A., Shi Y.P.,
RA   Wu R., Lin C.C., Coughlin S.R.;
RT   "Conserved structure and adjacent location of the thrombin receptor
RT   and protease-activated receptor 2 genes define a protease-activated
RT   receptor gene cluster.";
RL   Mol. Med. 2:349-357(1996).
RN   [3]
RP   SEQUENCE OF 1-74 FROM N.A.
RC   STRAIN=129/Sv;
RX   MEDLINE=96327649; PubMed=8678993;
RA   Xue J., Jenkins N.A., Gilbert D.J., Copeland N.G., Sadler J.E.;
RT   "Structure and localization of the thrombin receptor gene on mouse
RT   chromosome 13.";
RL   Mamm. Genome 7:625-626(1996).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L03529; AAA40438.1; -.
DR   EMBL; U36757; AAB38308.1; -.
DR   EMBL; U36756; AAB38308.1; JOINED.
DR   EMBL; U55075; AAB00198.1; -.
DR   EMBL; U55075; AAB00198.1; JOINED.
DR   HSSP; P34996; 1DDD.
DR   MGD; MGI:101802; F2r.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL        1      21        POTENTIAL.
```

```
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF028727; AAB84191.1; -.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL        1      26        POTENTIAL.
FT   PROPEP       27      41        REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN        42     425        PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       42     102        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    103     128        1 (POTENTIAL).
FT   DOMAIN      129     137        CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    138     157        2 (POTENTIAL).
FT   DOMAIN      158     176        EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    177     198        3 (POTENTIAL).
FT   DOMAIN      199     218        CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      219     239        4 (POTENTIAL).
FT   TRANSMEM    240     268        EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      269     288        5 (POTENTIAL).
FT   TRANSMEM    289     311        CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      312     334        6 (POTENTIAL).
FT   TRANSMEM    335     350        EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      351     374        7 (POTENTIAL).
FT   TRANSMEM    375     425        CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       57      60        ASP/GLU-RICH (ACIDIC).
FT   SITE         41      42        CLEAVAGE (BY THROMBIN).
FT   DISULFID    175     254        BY SIMILARITY.
FT   CARBOHYD     35      35        N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     62      62        N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     75      75        N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    250     250        N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    259     259        N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE   425 AA;  47253 MW;  118FC5FB3D4DE2C CRC64;
```

```
Query Match             19.1%;  Score 358.5;  DB 1;  Length 425;
Best Local Similarity   30.0%;  Pred. No.1.4e-18;
Matches  101;  Conservative  57;  Mismatches 158;  Indels  21;  Gaps  7;

Qy    25  SQTVKIFLRSLFYTALLVPSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP 84
              ::|: | ||  || :::|:|  : |:|| | :||::|  :::| |:|: : |:|
Db    98  SSWLTLFVPSV-YTGVFVVSLPVNIMAIVVFILKMKVKCPAVVYMLHLATADVLFVSVLP 156

Qy    85  GRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTA 144
              :: :: |: ||||:  ||| |  : :|:  ::| |||  ||:||
Db   157  FKISYYLSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTL 216

Qy   145  GRARLVCVAIWTLVLLQTMPLLLMPMKPLVGKLACMEYSSMESVLGLPLM------VLV 198
              | |::  |:| ||: :|| :::|           ::::  :: : |  |     :|
Db   217  GRASFTCLAIWALAIAGVVPLLLKEQTIQVPG---LNITTCHDVLNETLLEGYIAYFS 272

Qy   199  AF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS 257
             ||   : | | :::| : :|: : |:     :|::|:|:   :       :||   ::|||
Db   273  AFSAVFFFVPLIISTVCYVSIIRCLSSS-----TVANRSKKSRALFLSAAVFCIFIICFG 327

Qy   258  PYHLNIKQPMARGMLHLPSCAERRAFLLSLQATVALMNNNCGITPITIYFFASTHYRKWLL 317
             | ::|  ||||  :|| |:::  | | || : :|||||  ||:|   ||  ::| ||:
```

```
FT   PROPEP        22     41    REMOVED FOR RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CHAIN         42    430    PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN        42    107    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     108    133    1 (POTENTIAL).
FT   DOMAIN       134    142    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     143    162    2 (POTENTIAL).
FT   DOMAIN       163    181    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     182    203    3 (POTENTIAL).
FT   DOMAIN       204    223    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     224    244    4 (POTENTIAL).
FT   DOMAIN       245    273    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     274    293    5 (POTENTIAL).
FT   DOMAIN       294    316    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     317    339    6 (POTENTIAL).
FT   DOMAIN       340    354    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     355    379    7 (POTENTIAL).
FT   DOMAIN       380    430    CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        60     65    ASP/GLU-RICH (ACIDIC).
FT   DOMAIN        88     91    POLY-PRO.
FT   SITE          41     42    CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID     180    259    BY SIMILARITY.
FT   CARBOHYD      67     67    N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD      80     80    N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD     255    255    N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CONFLICT     162    162    F -> S (IN REF. 1).
FT   CONFLICT     189    189    G -> Y (IN REF. 1).
FT   CONFLICT     223    223    R -> G (IN REF. 1).
FT   CONFLICT     262    262    V -> L (IN REF. 1).
FT   CONFLICT     365    365    S -> T (IN REF. 1).
SQ   SEQUENCE     430 AA;  47790 MW;  395FD64FAE52C9BF CRC64;
```

```
Query Match      19.0%;  Score 357;  DB 1;  Length 430;
Best Local Similarity   28.6%;  Pred. No. 1.8e-18;
Matches  96;  Conservative  63;  Mismatches 163;  Indels 14;  Gaps  6;

Qy      2  IKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLIVFSALGNILALCLTCQKSRK  61
           | |  |||        ::       : :|: :| : | :  :|| |:   : |
Db     81  ISLPPH-TPPPPFISEDASGYLTSPWLTLFMPSV-YTIVFIVSLPLNVLAIAVFVLRMKV  138

Qy     62  INCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLM  121
           :|:||:|||:|||      :: :||  ||  |:|:| :|| |  | || | |||| ||
Db    139  KKPAVVVMLHLAMADVLFVSVLPFKISYYFSGDTWQFGSGMCRFATAAFYGNMYASIMLM  198

Qy    122  ACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTK--PLVGKLA  179
           |   |  |:| |   :  |       | :||  |  |  :|||   |  :    |
Db    199  TVISIDRFLAVVYPIQSLSWRTLGRANFTCVIWMVMAIMGVVPLLLKEQTTRVPGLNITT  258

Qy    180  CMEYSSMESVLGLPLMVLVAF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRH  238
           | :            :: |   |||  |  ||| :||       :|:
Db    259  CHDVLSENLMQGFYSYYFSAFSAIFFLVPLIVSTVCVTSIIRCLSSSA----VANRSKK  313

Qy    239  WRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNC  298
           |: |  ||       |     |: |||:  ||:|    |
Db    314  SRALFLSAAVFCIFIVCFGPTNV---LLIVHYLFLSDSPGTEAAYFAYLLCVCVSSVSC  369

Qy    299  GITPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSSS  334
           |  || |:|  : |:    ||:  |
Db    370  CIDPLIYYYASSECQRHLYSILCCKESSDPNSCNST  405
```

```
RESULT 6
PAR1_HUMAN
ID   PAR1_HUMAN      STANDARD;      PRT;     425 AA.
AC   P25116; Q96RF7; Q9BUN4;
DT   01-MAY-1992 (Rel. 22, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor)
DE   (Coagulation factor II receptor)
GN   F2R OR PAR1 OR TR OR CF2R.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91168254; PubMed=1672265;
RA   Vu T.-K.H., Hung D.T., Wheaton V.I., Coughlin S.R.;
RT   "Molecular cloning of a functional thrombin receptor reveals a novel
RT   proteolytic mechanism of receptor activation.";
RL   Cell 64:1057-1068(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,
RA   Nickerson D.A.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunarathne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=99178892; PubMed=10079109;
RA   Kahn M.L. Nakanishi-Matsui M., Shapiro M.J., Ishihara H.,
RA   Coughlin S.R.;
RT   "Protease-activated receptors 1 and 4 mediate activation of human
RT   platelets by thrombin.";
RL   J. Clin. Invest. 103:879-887(1999).
RN   [5]
RP   VARIANT GLY-166.
RX   MEDLINE=99318093; PubMed=10391209;
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RT   "Characterization of single-nucleotide polymorphisms in coding regions
RT   of human genes.";
RL   Nat. Genet. 22:231-238(1999).
RN   [6]
RP   ERRATUM.
RA   Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
RA   Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
RA   Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA   Lander E.S.;
RL   Nat. Genet. 23:373-373(1999).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis. May play a
CC       role in platelets activation and in vascular development.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: PLATELETS AND VASCULAR ENDOTHELIAL CELLS.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- PTM: Phosphorylated; probably mediating desensitisation prior to
CC       the uncoupling and internalization of the receptor.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
Query Match             18.8%;  Score 352.5;  DB 1;  Length 425;
Best Local Similarity   29.7%;  Pred. No. 3.8e-18;
Matches 100; Conservative  57; Mismatches 159;  Indels  21;  Gaps   7;

Qy     25  SQTVKIFLRSLFYTALLVFSALGNIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP  84
               :  : :|.:   |.|| |  :..||| |.::|.:|.|||.||:.||| |
Db     98  SSWLTLFVPSV-YTGVFVVSLPLNIMAIVVFILKMKVKKPAVVYMLHLATADVLFVSVLP 156

Qy     85  GRVVCYVLGSSWPFGKGLCRLTAFVLYTDTGGVVLMACVSVDHYPAVVCAFWGPRLRTA  144
               |.:: ||..::| .:|  .|: .: :|| .: ||.::.. ::.: :|.:|.|| |:
Db    157  FKISYYFSGSDWQFGSEICRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQSLSWRTL 216

Qy    145  GRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSMESVLGLPLM------VLV  198
               ||:|:|| .||.  .: :|  |  . .|.  : :|| ||:.. ::. |.|     |||
Db    217  GRASFTCIAIWALAIAGVVPLLLKEQTIQVPG---LNITTCHDVLNETLLEGYYAYYFS 272

Qy    199  AF-AIGFCGPVGIILSCYMKITWKLCCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS 257
               || ||.|| :|||:|.|.:|:| .| ||:|  .: :: ::.| ::..|.:  ||:..||
Db    273  AFSAVFFVPLIISTVCYVSIIRCLSSSA-----VANRSKKSRALFLSAAVFCIFIICFG 327

Qy    258  PYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPITPIYYFFASTHYRKWLL 317
               |.:||..  |.|:|:|| .| ||   .:|  |  |: ..| :| : :  ::.  .:|:|.||
Db    328  P--TNVLLIAHYSFLSHTSTTEAAYFAYLL--CVCVSISCCIDPLIYYASSECQRVVY 383

Qy    318  GILKLKGSSSSSSSSSSTPGKASSETPSITQARGSMF  354
               |.:| :|.:|||||||||  |.|:. ::. .|..:
Db    384  SILCCKESSDPSYNSSGQLMASKMDTCSSNLNNSIY 420

RESULT 7
V206_FOWPV
ID   V206_FOWPV     STANDARD;    PRT;   308 AA.
AC   Q9J529;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G-protein coupled receptor homolog FPV206.
GN   FPV206.
OS   Fowlpox virus (FPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Avipoxvirus.
OX   NCBI_TaxID=10261;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20193820; PubMed=10729156;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT   "The genome of fowlpox virus.";
RL   J. Virol. 74:3815-3831(2000).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF198100; AAF44550.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransmemBrane; Glycoprotein.
FT   DOMAIN         1     20       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      21     41       1 (POTENTIAL).
FT   DOMAIN        42     52       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      53     73       2 (POTENTIAL).
FT   DOMAIN        74     91       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      92    112       3 (POTENTIAL).
```

```
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M62424; AAA36743.1; -.
DR   EMBL; AF391809; AAK69768.1; -.
DR   EMBL; BC024464; AAH02464.1; -.
DR   PIR; A37912; A37912.
DR   PDB; 1NRN; 31-MAY-94.
DR   PDB; 1NRQ; 31-MAY-94.
DR   PDB; 1NRR; 31-MAY-94.
DR   Genew; HGNC:3537; F2R.
DR   MIM; 187930; -.
DR   GO; GO:0005794; C:Golgi apparatus; TAS.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005102; F:receptor binding activity; TAS.
DR   GO; GO:0015057; F:thrombin receptor activity; TAS.
DR   GO; GO:0006915; P:apoptosis; TAS.
DR   GO; GO:0007596; P:blood coagulation; TAS.
DR   GO; GO:0006919; P:caspase activation; TAS.
DR   GO; GO:0006928; P:cell motility; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0007125; P:invasive growth; TAS.
DR   GO; GO:0000074; P:regulation of cell cycle; TAS.
DR   GO; GO:0009611; P:response to wounding; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   GO; GO:0007262; P:STAT protein nuclear translocation; TAS.
DR   GO; GO:0007260; P:tyrosine phosphorylation of STAT protein; TAS.
DR   InterPro; IPR000276; 7tm_1; 1.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
DR   G-protein coupled receptor; TransmemBrane; Glycoprotein; Signal;
KW   Blood coagulation; Phosphorylation; Polymorphism; 3D-structure.
FT   SIGNAL         1     26       POTENTIAL.
FT   PROPEP        27     41       REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN         42    425       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN        42    102       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     103    128       1 (POTENTIAL).
FT   DOMAIN       129    137       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     138    157       2 (POTENTIAL).
FT   DOMAIN       158    176       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     177    198       3 (POTENTIAL).
FT   DOMAIN       199    218       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     219    239       4 (POTENTIAL).
FT   DOMAIN       240    268       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     269    288       5 (POTENTIAL).
FT   DOMAIN       289    311       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     312    334       6 (POTENTIAL).
FT   DOMAIN       335    350       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     351    374       7 (POTENTIAL).
FT   DOMAIN       375    425       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        57     60       ASP/GLU-RICH (ACIDIC).
FT   SITE          41     42       CLEAVAGE (BY THROMBIN).
FT   DISULFID     175    254       BY SIMILARITY.
FT   CARBOHYD      35     35       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD      62     62       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD      75     75       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD     250    250       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   CARBOHYD     259    259       N-LINKED (GLCNAC...)   (POTENTIAL).
FT   VARIANT      166    166       S -> G (IN dbSNP:5893)
FT                                 /FTId=VAR_014167.
FT   CONFLICT     238    238       L -> V (IN REF. 1).
FT   CONFLICT     335    335       A -> V (IN REF. 3).
FT   CONFLICT     364    364       C -> S (IN REF. 1).
SQ   SEQUENCE   425 AA;  47440 MW;  41B742A99EEC96AB CRC64;
```

```
FT   DOMAIN      113    133       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    134    154       4 (POTENTIAL).
FT   DOMAIN      155    180       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    181    201       5 (POTENTIAL).
FT   DOMAIN      202    226       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    227    247       6 (POTENTIAL).
FT   DOMAIN      248    270       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    271    291       7 (POTENTIAL).
FT   DOMAIN      292    308       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    89     167       BY SIMILARITY.
FT   CARBOHYD    2      2         N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .)  (POTENTIAL).
SQ   SEQUENCE    308 AA;  35375 MW;  5ABFE53EE1FEB28A CRC64;

Query Match    18.2%;  Score 341;  DB 1;  Length 308;
Best Local Similarity  29.8%;  Pred. No. 1.8e-17;
Matches   90;  Conservative  55;  Mismatches 129;  Indels  28;  Gaps   8;

Qy    34  SLFYTALLVFSA--LGNILALCL----TCQKSRKINCTGIYLVHLAVSDLLFTVALPGRV  87
              :|:|.||   :||||  .:     |||:||| ||||||||| ::
Db    19  SLAAVTILIFSSSLILNISALVIGFYTTAPGPMK-----MYLINLIVSDILFTVTLPLKI  73

Qy    88  VCYVLGSSWPFGKGLCRLTAFVLYTDYGGVVLMACVSVDHYPAVVCAHWGPRLRTAGRA  147
              |:|.:|:. ||||||||  .::||  :||:||  :||:|  |||||| |||:
Db    74  DYYYFFNWRWGEMACRIMSFLSYINTYVSINFMTWISVNRYAVTRPHKYNSRDNIMRT  133

Qy   148  RLVCVAIWTLVLLCTMPLLLMPMT-KPIVGKLACMEYSSMESVIGLPLMVLVAFA-IGFC 205
              |:.|.||||| ::.   || ||:  |:|||| ||| ::|:|:|  |||    .||
Db   134  KIACACTWIIILVPMSSILFVSTTSSDHETKIRCMEYNKVGDSMYLPPWVTIVMCFIGFV 193

Qy   206  GPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQ 265
              .:    ||.|  |:  :     |.||::|| : |||| ||| |||
Db   194  IPFAMMAISYSAVCYTVLSG-----ISKSTRSYRTCKLVACILTEFVICFLPYHASVIS 247

Qy   266  FMARGMLHLPS-----CAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL 320
              :.| :.       ||:  ::  :||||:  |:| |||:|||: ||||
Db   248  Y---MIHIITSKTVLCENVSYYQMLHATQCLMKLNCCMDPIIYLFVSSYKSKAKSNSI 303

Qy   321  KL  322
          ||
Db   304  KL  305


RESULT 8
PAR1_RAT
ID   PAR1_RAT    STANDARD;     PRT;    432 AA.
AC   P26824;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
GN   F2R OR PAR1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aortic smooth muscle;
RX   MEDLINE=92381002; PubMed=1324917;
RA   Zhong C., Hayzer D.J., Corsen M.A., Wick K., Runge M.S.;
RT   "Molecular cloning of the rat vascular smooth muscle thrombin
RT   receptor. Evidence for in vitro regulation by basic fibroblast growth
RT   factor.";
RL   J. Biol. Chem. 267:16975-16979(1992).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M81642; AAA42274.1; -.
DR   PIR; A43448; A43448.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL      1      21        POTENTIAL.
FT   PROPEP      22     45        REMOVED FOR RECEPTOR ACTIVATION (BY
FT                                SIMILARITY).
FT   CHAIN       46     432       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN      46     109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    110    135       1 (POTENTIAL).
FT   DOMAIN      136    144       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    145    164       2 (POTENTIAL).
FT   DOMAIN      165    183       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    184    205       3 (POTENTIAL).
FT   DOMAIN      206    225       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    226    246       4 (POTENTIAL).
FT   DOMAIN      247    275       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    276    295       5 (POTENTIAL).
FT   DOMAIN      296    318       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    319    341       6 (POTENTIAL).
FT   DOMAIN      342    357       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    358    381       7 (POTENTIAL).
FT   DOMAIN      382    432       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      64     68        ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      87     93        POLY-PRO.
FT   SITE        45     46        CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID    182    261       BY SIMILARITY.
FT   CARBOHYD    69     69        N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    82     82        N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    257    257       N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC. . .)  (POTENTIAL).
SQ   SEQUENCE    432 AA;  48280 MW;  DD032B97ABA4A605 CRC64;

Query Match    18.1%;  Score 339;  DB 1;  Length 432;
Best Local Similarity  29.0%;  Pred. No. 3.5e-17;
Matches   98;  Conservative  61;  Mismatches 153;  Indels  26;  Gaps   9;

Qy     6  PYFTPPTKIKTKIKDLN--VESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKIN  63
              :|   ||    |:  :..||:||:||||| |||||| ||||||||:|:||||
Db    87  PPMPPPPFIS---EDASGYLTSPWLTLFIPSV-YTFVFIVSLPLNILAIAVFVFRMKVKK 142

Qy    64  CTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYGGVVLMAC 123
              |||:|||||||||||||||||| ||.|:|||:|||| ||:|:|||||||||:||
Db   143  PAVVYMLHLAMADVLFVSVLPFKISYYFSGTDWQFSSGMCRFATAACYCNMYASIMLMTV 202

Qy   124  VSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEY 183
              :|::|.  ||||| |:  :.  :::|  |||:::|| ||:|||
Db   203  ISIDRFLAVVYPIQSLSWRTLGRANFTCVVIWVMAIMGVVPLLLKEQTTQVPG----LNI 258

Qy   184  SSMESVLGLPLM------VLVAF-AIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKG 236
              ||   :.||    :|||.|  |  ::|||  |:|:    |:|
Db   259  TTCHDVLNETLLHGFSYYFSAFSAIFFLVPLIISTVCYTSIIRCLSSSA----VANRS 313

Qy   237  RHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNM 296
              || :| |||  :||.|||| :||||
Db   314  KKSRALFLSAAVFCIFIVCFGPTNVLL---IVHYLLSDSPGTEYFAYLLCVCVTSVA 370

Qy   297  NCGITPIIYFFASTHYRKWLLGILKLGSSSSSSSS 334
              ||| ||:||||| ||||:|||| |:||::
Db   371  SC-IDFLIYYYASSECQKHLYSILCCRESSDSNSCNST 407
```

```
RESULT 9
PAR1_XENLA
ID   PAR1_XENLA     STANDARD;      PRT;   420 AA.
AC   P47749;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Proteinase activated receptor 1 precursor (PAR-1) (Thrombin receptor).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94195429; PubMed=8145852;
RA   Gerszten R.E., Chen J., Ishii M., Ishii K., Nanevicz T.,
RA   Turck C.W., Vu T.-K.H., Coughlin S.R.;
RT   "Specificity of the thrombin receptor for agonist peptide is defined
RT   by its extracellular surface.";
RL   Nature 368:648-651(1994).
CC   -!- FUNCTION: High affinity receptor for activated thrombin coupled to
CC       G proteins that stimulate phosphoinositide hydrolysis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: A proteolytic cleavage generates a new amino terminus that
CC       functions as a tethered ligand.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```
```
DR   EMBL; U09632; AAA18498.1; -.
DR   PIR; I51667; I51667.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; TransMembrane; Glycoprotein; Signal;
KW   Blood coagulation.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     42       REMOVED FOR RECEPTOR ACTIVATION.
FT   CHAIN        43    420       PROTEINASE ACTIVATED RECEPTOR 1.
FT   DOMAIN       43    101       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    102    127       1 (POTENTIAL).
FT   DOMAIN      128    136       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    137    156       2 (POTENTIAL).
FT   DOMAIN      157    175       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    176    197       3 (POTENTIAL).
FT   DOMAIN      198    217       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    218    238       4 (POTENTIAL).
FT   DOMAIN      239    267       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    268    287       5 (POTENTIAL).
FT   DOMAIN      288    310       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    311    333       6 (POTENTIAL).
FT   DOMAIN      334    345       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    346    369       7 (POTENTIAL).
FT   DOMAIN      370    420       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      295    300       POLY-SER.
FT   SITE         42     43       CLEAVAGE (BY THROMBIN).
FT   DISULFID    174    253       BY SIMILARITY.
FT   CARBOHYD     38     38       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     86     86       N-LINKED (GLCNAC...) (POTENTIAL).
FT   MUTAGEN      42     42       K->A: LACK OF ACTIVATION BY THROMBIN.
SQ   SEQUENCE   420 AA;  47435 MW;  D5163F56AFE12372 CRC64;
```

```
Query Match             17.7%;  Score 332.5;  DB 1;  Length 420;
Best Local Similarity   26.6%;  Pred. No. 9.7e-17;
Matches   92;  Conservative   74;  Mismatches 155;  Indels   25;  Gaps   8;

Qy   16 TKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIALCLTCQKSRKINCTGIYLVHLAVS  75
            :  :  :: |::| |: ||  |  | :  ||: : :|:|  :::||.||::
Db   88 TKEAEQYLSSQWLTKFVPSL-YTVFIVGLPLNLLAIIIFLFKMKVRKPAVVYMLNLAIA  146

Qy   76 DLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDYGGVYLMACVSVDHYPAVVCA  135
            |::| :::  |::  :||| |    |||:|  | | :|   | |:|| |:
Db  147 DVFFVSVLPFKIAYHLSGNDWLFGPGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYP  206

Qy  136 HWGPRLTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLM  195
         | ||::|: : | | |  : |  |:  |:::: ::|:||
Db  207 MESLSWRTMSRAYMACSFIWILSIASTIPLLVTEQTQ----KIPRLDITTCHDVLDLKDL  262

Qy  196 --VIVAPAIGFC------GPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLM  248
            ::    :         :  |::|     :::| |::|  : :  :
Db  263 KDFYIYPSSFCLLFFVPFITTICYIGIIRSLSSSSIENSC-----KKTRALFLAVVV  317

Qy  249 LVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFA  308
         |  :  :   :     |:|       || |::   : :| :  | :   ||::||::
Db  318 LCVFIICFGPTNV---LFLTH---YLQEANEFLYFAYILSACVG--SVSCCLDPLIYYYA  369

Qy  309 STHYRKWLLGILKLKGSSSSSSSSSSSSSTPGKASSETPSITQARGSMF  354
             :   : :::| :|::::::::::::
Db  370 SSQCQRYIYSLLCCRKVSEPGSSTGQLMSTAMKNDNCSTNAKSSIY  415
```

```
RESULT 10
P2Y8_XENLA
ID   P2Y8_XENLA     STANDARD;      PRT;   537 AA.
AC   P79928;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   P2Y purinoceptor 8 (P2Y8).
GN   P2RY8.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neural plate;
RX   MEDLINE=97284734; PubMed=9139711;
RA   Bogdanov Y.D., Dale L., King B.F., Whittock N., Burnstock G.;
RT   "Early expression of a novel nucleotide receptor in the neural plate
RT   of Xenopus embryos.";
RL   J. Biol. Chem. 272:12583-12590(1997).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR ATP, UTP, CTP, GTP AND ITP.
CC       THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
CC       ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC       MAY PLAY A KEY ROLE IN THE EARLY DEVELOPMENT OF NEURAL TISSUE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```
```
DR   EMBL; X99953; CAA68213.1; -.
DR   HSSP; P34996; 1DDD.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
```

```
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     49       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     50     70       1 (POTENTIAL).
FT   DOMAIN       71     79       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     80    100       2 (POTENTIAL).
FT   DOMAIN      101    118       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    119    139       3 (POTENTIAL).
FT   DOMAIN      140    161       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    162    182       4 (POTENTIAL).
FT   DOMAIN      183    210       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    211    231       5 (POTENTIAL).
FT   DOMAIN      232    254       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    255    275       6 (POTENTIAL).
FT   DOMAIN      276    292       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    293    316       7 (POTENTIAL).
FT   DOMAIN      317    537       CYTOPLASMIC (POTENTIAL).
FT   DISULFID    116    193       BY SIMILARITY.
FT   CARBOHYD     26     26       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     29     29       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   537 AA;  62024 MW;  B2CF24812F3C19F2 CRC64;

Query Match          17.7%;  Score 332;  DB 1;  Length 537;
Best Local Similarity  25.1%;  Pred. No. 1.3e-16;
Matches  97;  Conservative  70;  Mismatches  162;  Indels  58;  Gaps  10;

Qy    6  PYFTPPTKIKTKI----KDLNVESQTVKIFLRSLFTTALLVFSALGNILACLTCQKSRK   61
            |  |  : ||  :|:|   :|:|  | |:   | :||  || ||::|:
Db   17  PYL--PMKLLMNLITNDTEDICVFDEGFKFLLLPVSYSAVFMVGLPLNIIAAMWIFIAKMRP   74

Qy   62  INCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPPGKGLCRLTAFVLYTDTYGGVYLM  121
             |:||:| |: |||  | :|: | ||  ||   :||| |||  | | :| |: |:||
Db   75  WNPTTVYMFNLALSDTLYVLSLPTLVYYADKNNWPFGEVLCKLVRFLFYANLYSSILFL   134

Qy  122  ACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC-  180
            ::||          :|  :      | |:    :|  | ::| | : | ||  ::
Db  135  TCISVHRYRGVCHPITSLRRVNAKHAYVICALVWLSVTLCLVPNLIFVTVSPKVKNTICH  194

Qy  181  ------MEYSS-MESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCS---  224
                  :|::    :| | | :: :    |  |   | ::  |::    |
Db  195  DTTRPEDFARVVEYSTAIMCLLFGIPCL--------IIAGCYGLMTRELMKPIV  240

Qy  225  TAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFL  284
             :  |  :||:  |||  |||:| |:|::||||    |  :|| | ||  |:  | |
Db  241  SGNQQTLPSYKKRSIK---TIIFVMIAFAICFMPFHFITRTLYYARLLGI-KCYALNVIN  296

Qy  285  LSLQATVALMNNNCGITPIIYFPFASTHYRKWLLGIKLKLKGS----------SSSSSS  331
             |    |     |  |||  :| ||:  |:| ||  |: |||          ||||||
Db  297  VTYKVTRPLASANSCIDPILYFLANDRYRRRLRTVRRRSSVPNRRCMHTNHPQTEPHMT  356

Qy  332  SSTPGKASSETPSITQARGSMFPLAEH  358
            :  |::  |    |: | |  ||:
Db  357  AGPLPVISAEEIPS----NGSMVRDEN  379
```

RESULT 11
P2Y6_HUMAN

```
ID   P2Y6_HUMAN     STANDARD;      PRT;   328 AA.
AC   Q15077; Q15754;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   P2Y purinoceptor 6 (P2Y6).
GN   P2RY6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Communi D., Parmentier M., Boeynaems J.M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97432828; PubMed=9285708;
RA   Somers G.R., Hammet F., Woollatt E., Richards R.I., Southey M.C.,
RA   Venter D.J.;
RT   "Chromosomal localization of the human P2y6 purinoceptor gene and
RT   phylogenetic analysis of the P2y purinoceptor family.";
RL   Genomics 44:127-130(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97415792; PubMed=9268704;
RA   Maier R., Glatz A., Mosbacher J., Bilbe G.;
RT   "Cloning of P2Y6 cDNAs and identification of a pseudogene: comparison
RT   of P2y receptor subtype expression in bone and brain tissues.";
RL   Biochem. Biophys. Res. Commun. 237:297-302(1997).
RN   [4]
RP   ERRATUM.
RX   MEDLINE=98069816; PubMed=9412455;
RA   Maier R., Glatz A., Mosbacher J., Bilbe G.;
RL   Biochem. Biophys. Res. Commun. 240:298-302(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA   Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA   Yu J., Han L.H.;
RT   "Novel human cDNA clones with function of inhibiting cancer cell
RT   growth.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
CC       OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
CC       PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
```

```
DR   EMBL; X97058; CAA65770.1; -.
DR   EMBL; U52464; AAB03572.1; -.
DR   EMBL; AF007892; AAB80713.1; -.
DR   EMBL; AF007891; AAB80712.1; -.
```

```
DR   EMBL; AF218005; AAG17247.1; -.
DR   EMBL; AF498920; AAM18129.1; -.
DR   EMBL; BC000571; AAH00571.1; -.
DR   EMBL; AC009391; AAH09391.1; -.
DR   PIR; JC4800; JC4800.
DR   HSSP; P34996; 1DDD.
DR   Genew; HGNC:8543; P2RY6.
DR   MIM; 602451; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0007200; P:G-protein signaling, coupled to IP3 second ... .; TAS.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     27       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     28     48       1 (POTENTIAL).
FT   DOMAIN       49     62       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     63     83       2 (POTENTIAL).
FT   DOMAIN       84    101       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    102    122       3 (POTENTIAL).
FT   DOMAIN      123    144       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    145    165       4 (POTENTIAL).
FT   DOMAIN      166    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    215       5 (POTENTIAL).
FT   DOMAIN      216    236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    237    257       6 (POTENTIAL).
FT   DOMAIN      258    280       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    281    303       7 (POTENTIAL).
FT   DOMAIN      304    328       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD      5      5       N-LINKED (GLCNAC...) (POTENTIAL).
FT   DISULFID     99    177       BY SIMILARITY.
FT   CONFLICT      2      3       MISSING (IN REF. 2).
SQ   SEQUENCE   328 AA;  36429 MW;  AAD6C55A43818107 CRC64;

Query Match             17.5%;  Score 329;  DB 1;  Length 328;
Best Local Similarity   28.0%;  Pred. No. 1.4e-16;
Matches   89;  Conservative  54;  Mismatches 141;  Indels  34;  Gaps  11;

Qy        10 PPTKIKTKIKDLNVSQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYL 69
                 |||    ||| :|:| :|:|   ||:|   :|| | |:|
Db        14 PPTTC------VYRENFKQLLLPPVYSAVLAAGLPLNICVITQICTSRRALTRTAVYT 65

Qy        70 VHLAVSDLLFTVALPGRVVCVVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHY 129
             ::|||:|||: :||| :|: || | |  |  ||: :|  ||:|||: :||:|:: |::
Db        66 LNLALDLLYACSLPLLIYNYAQGDHWPFGDFACRLVRFLFYANLHGSILFLTCISFQRY 125

Qy       130 PAVVCAHWGPRLRTAGR--ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYS--- 184
             | |:| :|| ||| :|    |:|:|| ||::|   :||||||  |:| |:|:
Db       126 LG-ICHPLAPWHKRGGRRAAWLVCVAVWLAVTTQCLPTAIFAATGIQRNRTVCYDLSPPA 184

Qy       185 --SMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTARED---PVT-SRKGR 237
               :: |:|  | |:   |  :|   |||| :|  :| ||| :         |:| |: |
Db       185 LATHYMPYGMALTV-----IGFLLFPAALLACYLLACRLC---RQDGPAEPVAQERRGK 236

Qy       238 HWRGCLLTLLMLVAVVVCFSPYHLNIKQFMA-RGMLHLPSCAERRAFLLSLQATVALMM 296
             | :| ||: |: | |:|||| ::|  |     |::: ||:  |  : |: |: |:| |
Db       237 AAR---MAVVVAAAFAISFLPFHITKTAYLAVRSTPGVP-CTVLEAFAAAYKGTRPFASA 292

Qy       297 NCGITPIIYFFASTHYRK 314
             | |   ||||   : : |
Db       293 NSVLDPILFYPTQKKFRR 310
```

```
RESULT 12
CKR8_MOUSE
ID   CKR8_MOUSE     STANDARD;      PRT;    353 AA.
AC   P56484;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   C-C chemokine receptor type 8 (C-C CKR-8) (CC-CKR-8) (CCR-8).
```

```
Qy    65  TGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACV 124
              |||::|.||::  |||||  ::|||  :  :|||:|  |:::||  ||:  ::::  |:
Db    68  TDIYLLNLAASDLLFVLSIPFQT--HNLLDQWVFGTAMCKVVSGLYYIGFFSSMFFITLM 125

Qy   125  SVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYS 184
              ||:|::|:  ||:  |:|     :| ::|  :|  :: |||:|  |||  || ::: :
Db   126  SVDRYLAIVHAVVAIKVRTASVGTALSLITVWLAAVTATIPLMVFYQVASEDGMLQCFQFY 185

Qy   185  SMESVLGLPLMLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCL- 243
              |:|  ||:  ||.||: |||   |:|:|  :|:| :|:|
Db   186  EEQSLRWKLFTHFEINALGLLLPFAILLFCYVRIIQQL---------------RGCLN 228

Qy   244  ------LTILMLVAVVVCFSPYHLNIKQFMARGMLH----LPSCAERRAFLLSLQATV 291
                    :|::|  |  ::  :.|  |:|  :||||  |:
Db   229  HNRTRAIKLVLTVVIVSLLFWVPF--NVALFLT--SLHDLHILDGCATRQRLALAIHVTE 284

Qy   292  ALMNNNCGITPIIYFFASTHYRKWLL----------------GILKLKGSSSSSSS 331
              :|  || :||||:||||||||  |:             :|::  ||||||||||
Db   285  VISFTHCCVNPVIYAFIGEKFKKGHLMDVFQKSCSHIFLYLGROMPVGALERQLSSNQRSS 344

Qy   332  SST 335
              |||
Db   345  HSST 348
```

RESULT 13
CKR8_HUMAN

ID   CKR8_HUMAN   STANDARD;   PRT;   355 AA.
AC   P51685;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   C-C chemokine receptor type 8 (C-C CKR-8) (CC-CKR-8) (CCR-8) (CKR-L1) (GPR-CY6)
DE   (GPRCY6) (Chemokine receptor-like 1) (CKR-L1) (TER1) (CMKBRL2) (CC-
DE   chemokine receptor CHEMR1).
GN   CCR8 OR CMKBR8 OR CKRL1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo..
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97351133; PubMed=9207005;
RA   Tiffany H.L., Lautens L.L., Gao J.-L., Pease J., Locati M.,
RA   Combadiere C., Modi W., Bonner T.I., Murphy P.M.;
RT   "Identification of CCR8: a human monocyte and thymus receptor for the
RT   CC chemokine I-309.";
RL   J. Exp. Med. 186:165-170(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98129363; PubMed=9469461;
RA   Goya I., Gutierrez J., Varona R., Kremer L., Zaballos A., Marquez G.;
RT   "Identification of CCR8 as the specific receptor for the human beta-
RT   chemokine I-309: cloning and molecular characterization of murine
RT   CCR8 as the receptor for TCA-3.";
RL   J. Immunol. 160:1975-1981(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97040707; PubMed=8886020;
RA   Zaballos A., Varona R., Gutierrez J., Lind P., Marquez G.;
RT   "Molecular cloning and RNA expression of two new human chemokine
RT   receptor-like genes.";
RL   Biochem. Biophys. Res. Commun. 227:846-853(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97131825; PubMed=8977299;
RA   Samson M., Stordeur P., Labbe O., Soularue P., Vassart G.,
RA   Parmentier M.;
RT   "Molecular cloning and chromosomal mapping of a novel human gene,
RT   ChemR1, expressed in T lymphocytes and polymorphonuclear cells and
RT   encoding a putative chemokine receptor.";
RL   Eur. J. Immunol. 26:3021-3028(1996).
RN   [5]

RP   SEQUENCE FROM N.A.
RA   Nakajima T., Yoshida R., Harada S.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   LIGAND BINDING.
RX   MEDLINE=98180363; PubMed=9521068;
RA   Bernardini G., Hedrick J., Sozzani S., Luini W., Spinetti G.,
RA   Weiss M., Menon S., Zlotnik A., Mantovani A., Santoni A.,
RA   Napolitano M.;
RT   "Identification of the CC chemokines TARC and macrophage inflammatory
RT   protein-1 beta as novel functional ligands for the CCR8 receptor.";
RL   Eur. J. Immunol. 28:582-588(1998).
CC   -!- FUNCTION: Receptor for the chemokines SCYA1/I-309, SCYA4/MIP-1-
CC   beta and SCYA17/TARC. May regulate monocyte chemotaxis and thymic
CC   cell line apoptosis. Alternative coreceptor with CD4 for HIV-1
CC   infection.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   --
DR   EMBL; U45983; AAB61962.1; -.
DR   EMBL; U62556; AAB05542.1; -.
DR   EMBL; Z79782; CAB02142.1; -.
DR   EMBL; Y08456; CAA69712.1; -.
DR   EMBL; D49919; BAA23387.1; -.
DR   EMBL; AF005210; AAB62547.1; -.
DR   PIR; JC5067; JC5067.
DR   Genew; HGNC:1609; CCR8.
DR   MIM; 601834; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0015026; F:coreceptor activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0006935; P:chemotaxis; TAS.
DR   GO; GO:0007204; P:cytosolic calcium ion concentration elevation; TAS.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin... .; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   InterPro; IPR004068; CC_chemokine8.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR01530; CHEMOKINER8.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN   1   35   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   36   63   1 (POTENTIAL).
FT   DOMAIN   64   73   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   74   93   2 (POTENTIAL).
FT   DOMAIN   94   107   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   108   129   3 (POTENTIAL).
FT   DOMAIN   130   146   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   147   171   4 (POTENTIAL).
FT   DOMAIN   172   202   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   203   222   5 (POTENTIAL).
FT   DOMAIN   223   238   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   239   263   6 (POTENTIAL).
FT   DOMAIN   264   280   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   281   304   7 (POTENTIAL).
FT   DOMAIN   305   355   CYTOPLASMIC (POTENTIAL).
FT   DISULFID   106   183   BY SIMILARITY.
SQ   SEQUENCE   355 AA; 40844 MW; BC14A153CF695361 CRC64;

Query Match   17.3%; Score 324; DB 1; Length 355;
Best Local Similarity   26.4%; Pred. No. 3.3e-16;
Matches   89; Conservative   65; Mismatches 145; Indels   38; Gaps   8;
```

```
Qy   26  QTVKIFLRSLFYTALLVFSALGN--ILALCLTCCQKSRKINCTGIYLVHLAVSDLLFTVAL   83
             ::||  ||:| |||  || :|:   |:||||  :||||||
Db   31  QTNGKLLLAVFYCLLFVFSLLGNSLVIVLVVCKKLRSI--TDVYLNLALSDLLFVFSF     88

Qy   84  PGRVVCYVLGSSWPPGKGLCRLITAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT  143
             |: :|  |||  ||| |::  ||:  :|||  ||  :|||  ||
Db   89  PFQT--YYLLDQWVFGTVMCKVVSGPFYYIGFYSSMFFITLMSVDRYLAVVHAVYALKVRT  146

Qy  144  AGRARLVCVAIWTLVLLQTMPLLIMPMTKPLVGKLACMEYSSMESVLGLPLMVIVAPAIG   203
             |:| :|:|  |:  :|:||   |||  : |
Db  147  IRMGTTLCLAVWLTAIMATIPLLVFYQVASEDGVLQCYSFYNQQT---LKWKIFTNFKWN   203

Qy  204  FCG---PVGIILSCYMKITWKL--CSTAEDPVTSRKGRHWRGCLLTLLMLVAVVVCFSP   258
              |   ||:  |  :|| : :   |  |||||  || :||||  | |:|
Db  204  ILGLLIPFTIFMFCYIKILHQLKRCQNHNKT-------KAIRLVLIVIASLLFWVP      253

Qy  259  YHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNMCGITPIIYFFASTHYRKWI--  316
              :|  :: |: :|    ||| ||||| :|  |||  ||||||| :|
Db  254  FNVVLFLTSLHSMHILDGCISQQLTYATHVTEIISFTHCCVNPVIYAFVGEKFKKHLSE  313

Qy  317  ----LGIILKLKGSSSSSSSSSSTPGKASS                              341
              |  :|||||||  |:|
Db  314  IFQKSCSQIFNYLGRQMPRESCEKSSSSCCQHSSSRSSS                     350


RESULT 14
CLT1_PIG
ID   CLT1_PIG      STANDARD;      PRT;   340 AA.
AC   Q95N02;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cysteinyl leukotriene receptor 1 (CysLTR1).
GN   CYSLTR1 OR CYSLT1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Takasaki J., Kamohara M., Saito T., Matsumoto M., Matsumoto S.,
RA   Ohishi T., Soga T., Matsushime H., Furuichi K.;
RT   "Characterization of cloned rat and porcine cysteinyl leukotriene
RT   receptors.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Receptor for cysteinyl leukotrienes mediating
CC       constriction of the microvascular smooth muscle during an
CC       inflammatory response. This response is mediated via a G-protein
CC       that activates a phosphatidylinositol-calcium second messenger
CC       system (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AB052686; BAB60826.1; -.
DR   InterPro; IPR0004071; Cysleuk_receptor.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; -.
DR   PRINTS; PR01533; CYSLTRECPTR.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein.
FT   DOMAIN        1     31    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     32     52    1 (POTENTIAL).
FT   DOMAIN       53     60    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     61     81    2 (POTENTIAL).
FT   DOMAIN       82    109    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    110    130    3 (POTENTIAL).
FT   DOMAIN      131    144    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    145    165    4 (POTENTIAL).
FT   DOMAIN      166    196    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    197    217    5 (POTENTIAL).
FT   DOMAIN      218    233    CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    234    254    6 (POTENTIAL).
FT   DOMAIN      255    279    EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    280    300    7 (POTENTIAL).
FT   DOMAIN      301    340    CYTOPLASMIC (POTENTIAL).
FT   DISULFID     99    176    BY SIMILARITY.
FT   CARBOHYD      6      6    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     18     18    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    172    172    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    265    265    N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   340 AA;  38986 MW;  54F9372A12ICE413 CRC64;

Query Match             17.2%;  Score 323.5;  DB 1;  Length 340;
Best Local Similarity   27.5%;
Matches   90;  Conservative  62;  Mismatches  156;  Indels  19;  Gaps   6;  Pred. No. 3.5e-16;


Qy   22  NVESQTVKIF--LRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLL   78
             ::|   ||   :|  ||  |  ||  :|:| ||| : :||| |::|:|||| :|||
Db   15  NTCNDTIDDFRNQVYSTLYSMTVVGFFGNGFVLYVLIKTYHEKSAYQVYMNLAVADLL   74

Qy   79  FTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWG  138
             :||     |||  |||| :  ||  |:  :|: :|| |  |:| :|| || :|
Db   75  CVCTLPLRVVYYVHKGIWLFGDFLCRLSTYALYVNLYCSIFFMTAMSFFRCLAIVFPVQN  134

Qy  139  PRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACME-YSSMESVLGLPLMVL  197
             :|:  ||:|:|||  || :|::  :  | |   |  |: : |   ::  |
Db  135  INLITHKKAKIVCIAIWIFVILTSSPFLMSTSYKDBKNNTKCFEPPOXNQAKYHVLVLHY  194

Qy  198  VAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS  257
             :: ||  :  ||| :: :|:   ::||||    |:  :||||  | |:|
Db  195  VSLFVGFIIPFFVIIVCYTMILTLLKNSMKKNISSRK---KAIGMIVVTAAFLISFM    250

Qy  258  PYHLNIK---QFMARGMLHLPSCAERRAFLLSLQATVALMNNMCGITPIIYFFASTHYRK  314
             ||| |    :|  :|  |:|||| |:  |  : :|  |  ||||||| :|
Db  251  PYHIQRTIHLHFLHNDTKGHCDSVLRMQK--SVXITLSLAASNCCFDPLLYFSGGNFRE  307

Qy  315  WLLGILKLKGSSSSSSSSSSSSTPGKASS                                341
             :|  :|||:|
Db  308  GLSTFRK----HSLSTMTYVPKKKTS                                  329


RESULT 15
CKR4_HUMAN
ID   CKR4_HUMAN    STANDARD;      PRT;   360 AA.
AC   P51679; Q9ULY6; Q9ULY7;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C-C chemokine receptor type 4 (C-C CKR-4) (CC-CKR-4) (CCR-4) (CCR4)
DE   (K5-5).
GN   CCR4 OR CMKBR4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=95370289; PubMed=7642634;
RA   Power C.A., Meyer A., Nemeth K., Bacon K.B., Hoogewerf A.J.,
RA   Proudfoot A.E.I., Wells T.N.C.;
RT   "Molecular cloning and functional expression of a novel CC chemokine
RT   receptor cDNA from a human basophilic cell line.";
RL   J. Biol. Chem. 270:19495-19500(1995).
RN   [2]
RP   SEQUENCE FROM N.A. AND VARIANTS VAL-130 AND SER-178.
```

```
RX   MEDLINE=21040311; PubMed=11196669;
RA   Kato H., Tsuchiya N., Izumi S., Miyamasu M., Nakajima T., Kawasaki H.,
RA   Hirai K., Tokunaga K.;
RT   "New variations of human CC-chemokine receptors CCR3 and CCR4.";
RL   Genes Immun. 1:97-104(1999).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=97313486; PubMed=9169480;
RA   Imai T., Baba M., Nishimura M., Kakizaki M., Takagi S., Yoshie O.;
RT   "The T cell-directed CC chemokine TARC is a highly specific
RT   biological ligand for CC chemokine receptor 4.";
RL   J. Biol. Chem. 272:15036-15042(1997).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=98104168; PubMed=9430724;
RA   Imai T., Chantry D., Raport C.J., Wood C.L., Nishimura M., Godiska R.,
RA   Yoshie O., Gray P.W.;
RT   "Macrophage-derived chemokine is a functional ligand for the CC
RT   chemokine receptor 4.";
RL   J. Biol. Chem. 273:1764-1768 (1998).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=99394604; PubMed=10466728;
RA   Campbell J.J., Haraldsen G., Pan J., Rottman J., Qin S., Ponath P.,
RA   Andrew D.P., Warnke R., Ruffing N., Kassam N., Wu L., Butcher E.C.;
RT   "The chemokine receptor CCR4 in vascular recognition by cutaneous but
RT   not intestinal memory T cells.";
RL   Nature 400:776-780(1999).
RN   [6]
RP   FUNCTION, TISSUE SPECIFICITY, AND PHOSPHORYLATION.
RX   MEDLINE=20219238; PubMed=10754297;
RA   Inngjerdingen M., Damaj B., Maghazachi A.A.;
RT   "Human NK cells express CC chemokine receptors 4 and 8 and respond to
RT   thymus and activation-regulated chemokine, macrophage-derived
RT   chemokine, and I-309.";
RL   J. Immuncl. 164:4048-4054(2000).
CC   -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR THE C-C TYPE CHEMOKINES
CC       TARC/SCYA17 AND MDC/SCYA22. THE ACTIVITY OF THIS RECEPTOR IS
CC       MEDIATED BY G(I) PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-
CC       CALCIUM SECOND MESSENGER SYSTEM. CAN FUNCTION AS A CHEMOATTRACTANT
CC       HOMING RECEPTOR ON CIRCULATING MEMORY LYMPHOCYTES AND AS A
CC       CORECEPTOR FOR SOME PRIMARY HIV-2 ISOLATES. IN THE CNS, COULD
CC       MEDIATE HIPPOCAMPAL-NEURON SURVIVAL.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE THYMUS, IN
CC       PERIPHERAL BLOOD LEUKOCYTES, INCLUDING T CELLS, MOSTLY CD4+ CELLS,
CC       AND BASOPHILS, AND IN PLATELETS; AT LOWER LEVELS, IN THE SPLEEN
CC       AND IN MONOCYTES. DETECTED ALSO IN MACROPHAGES, IL-2-ACTIVATED
CC       NATURAL KILLER CELLS AND SKIN-HOMING MEMORY T CELLS, MOSTLY THE
CC       ONES EXPRESSING THE CUTANEOUS LYMPHOCYTE ANTIGEN (CLA). EXPRESSED
CC       IN BRAIN MICROVASCULAR AND CORONARY ARTERY ENDOTHELIAL CELLS.
CC   -!- PTM: IN NATURAL KILLER CELLS, SCYA22 BINDING INDUCES
CC       PHOSPHORYLATION ON YET UNDEFINED SER/THR RESIDUES, MOST PROBABLY
CC       BY BETA-ADRENERGIC RECEPTOR KINASES 1 AND 2.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X85740; CAA59743.1; -.
DR   EMBL; AB023888; BAA86965.1; -.
DR   EMBL; AB023889; BAA86966.1; -.
DR   EMBL; AB023890; BAA86967.1; -.
DR   EMBL; AB023891; BAA86968.1; -.
DR   EMBL; AB023892; BAA86969.1; -.
DR   PIR; A57160; A57160.
DR   Genew; HGNC:1605; CCR4.
DR   MIM; 604836; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004950; F:chemokine receptor activity; TAS.
DR   GO; GO:0006935; P:chemotaxis; TAS.
DR   GO; GO:0007204; P:cytosolic calcium ion concentration elevation; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   GO; GO:0006954; P:inflammatory response; TAS.
DR   InterPro; IPR000276; GPCR_Rhodopsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Polymorphism.
FT   DOMAIN        1     39       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     40     67       1 (POTENTIAL).
FT   DOMAIN       68     77       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     78     98       2 (POTENTIAL).
FT   DOMAIN       99    111       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    112    133       3 (POTENTIAL).
FT   DOMAIN      134    150       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    151    175       4 (POTENTIAL).
FT   DOMAIN      176    206       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    207    226       5 (POTENTIAL).
FT   DOMAIN      227    242       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    243    267       6 (POTENTIAL).
FT   DOMAIN      268    284       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    285    308       7 (POTENTIAL).
FT   DOMAIN      309    360       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    183    183       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC...) (POTENTIAL).
FT   DISULFID    110    187       BY SIMILARITY.
FT   VARIANT     130    130       L -> V.
FT                                /FTId=VAR_010669.
FT   VARIANT     178    178       C -> S.
FT                                /FTId=VAR_010670.
SQ   SEQUENCE   360 AA;  41402 MW;  51EBE12ADIFAFABF CRC64;
```

```
Query Match             17.2%;  Score 322.5;  DB 1;  Length 360;
Best Local Similarity   28.3%;  Pred. No. 4.3e-16;
Matches   93;  Conservative  58;  Mismatches  131;  Indels  47;  Gaps     9;

Qy     29  KIFLRSLFYTALLVFSAIGNIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP--GR  86
               ::|:|||||    |||| ||| :: :  |   :|||::||| |||||:||:||
Db     39  ELFLPPI-YSLVFVFGLLGNSVVVLVLFKYKYKLLRSMTDVYLLNLAISDLLFVFSLPWG-  96

Qy     87  VVCYVLGSSWPFGKGLCRLTAFLVYLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGR  146
              ::|  :||:  ||:|  |::: :||::  :|||:|||||||::|:| |:
Db     97  --YYAADQWVFGLGLCKMISMYLVCFYSGIFFVMLMSIDRYLAIVHAVFSLRARTLTY  153

Qy    147  ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCG  206
              :|||   :|   |  :| ||||||: |||::: :|::|||  :|:|:||: ::
Db    154  GVTSLAATWSVAVFASLPGFLFSTCYTERNHTYCKTKYSLNSTTWKVLSSLEINILGLVI  213

Qy    207  PVGIILSCYMKI--TWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCF----SPYH  260
              :||:||  :    :||  :: :::  ::|  ||
Db    214  PLGIMLFCYSMIIRTLQHCKNEKGNKAVK-----------MIFAVVVLFLGFWTPY-  258

Qy    261  LNIKQFMAR--GMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLL-  317
              |:||| :     : :::  :|                ::|  ||:||||  |  :::|
Db    259  -NIVLFLTIVELEVLQDCTFERYLDYAIQATETLAFVHCCLNPIIYFFLGEKFRKYILQ  317

Qy    318  -------------GILKLKGSSSSSSS  331
                          ||:| ::: ||:
Db    318  LFKTCRGLFVLCQYCGLLQIYSADTPSSS  346

Search completed: February 19, 2004, 19:33:29
Job time : 18 secs
```

```
          OM protein - protein search, using sw model

                    Copyright (c) 1993 - 2004 Compugen Ltd.
          GenCore version 5.1.6

Run on:   February 19, 2004, 19:29:05 ; Search time 41 Seconds
                                         (without alignments)
                                      2265.828 Million cell updates/sec

Title:          US-10-041-615-34
Perfect score:  1876
Sequence:       1 MIKLGPYFTPPTKIKTKIKD........SETPSITQARGSMFLAEHVV 360

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:        830525

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :      SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*
                15: sp_rvirus:*
                16: sp_bacteriap:*
                17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	387.5	20.7	339	4	Q8N5S7	Q8N5S7 homo sapien
2	332.5	17.7	361	11	Q8VHP3	Q8VHP3 cavia porce
3	324	17.3	355	4	Q9BYX5	Q9BYX5 homo sapien
4	323	17.2	337	4	Q8IV19	Q8IV19 homo sapien
5	316.5	16.9	372	4	Q9H1C0	Q9H1C0 homo sapien
6	314.5	16.8	360	11	Q91ZH4	Q91ZH4 rattus norv
7	313.5	16.7	355	6	Q8HZN4	Q8HZN4 cercopithec
8	310.5	16.6	355	6	Q8HZN5	Q8HZN5 macaca mula
9	309.5	16.6	355	6	Q8HZN6	Q8HZN6 pongo pygma
10	309	16.5	355	6	Q9MYJ8	Q9MYJ8 callithrix
11	308.5	16.4	360	6	Q8MJW8	Q8MJW8 canis famil
12	306.5	16.3	355	4	Q8IUZ1	Q8IUZ1 homo sapien
13	306.5	16.3	355	6	Q8HZN8	Q8HZN8 pan troglod
14	306.5	16.3	362	4	Q8NE10	Q8NE10 homo sapien
15	305	16.3	352	11	Q8BY68	Q8BY68 mus musculu
16	304.5	16.2	355	6	Q8HZN7	Q8HZN7 gorilla gor
17	304	16.2	342	13	Q93239	Q93239 cyprinus ca
18	304	16.2	374	13	O57466	O57466 meleagris g
19	303.5	16.2	344	11	Q8BMC0	Q8BMC0 mus musculu
20	300.5	16.0	355	6	Q8HZN3	Q8HZN3 papio hamad
21	300	16.0	367	11	Q9QWN6	Q9QWN6 mus musculu
22	299.5	15.8	357	13	Q9DE05	Q9DE05 mus musculu
23	296	15.8	328	11	Q9ERK9	Q9ERK9 mus musculu
24	294.5	15.7	331	11	Q8K1Z6	Q8K1Z6 mus musculu
25	294.5	15.7	367	11	Q9JII9	Q9JII9 rattus norv
26	293.5	15.6	370	11	Q8BKK1	Q8BKK1 mus musculu
27	290.5	15.5	362	11	Q9JLZ0	Q9JLZ0 rattus norv
28	290.5	15.5	362	11	Q91WI0	Q91WI0 mus musculu
29	289.5	15.4	309	11	Q8R528	Q8R528 mus musculu
30	289.5	15.4	370	11	Q8BLG2	Q8BLG2 mus musculu
31	288.5	15.4	350	4	Q8N6T6	Q8N6T6 homo sapien
32	288.5	15.4	375	11	Q8BYI1	Q8BYI1 mus musculu
33	288	15.4	365	11	Q8BUD0	Q8BUD0 mus musculu
34	288	15.4	372	13	O93237	O93237 cyprinus ca
35	287	15.3	346	4	Q9BXC0	Q9BXC0 homo sapien
36	287	15.3	360	13	Q8QFR6	Q8QFR6 xenopus lae
37	286.5	15.3	353	13	Q9PTF7	Q9PTF7 brachydanio
38	286	15.2	377	13	Q98U14	Q98U14 brachydanio
39	284.5	15.2	359	11	Q8K3M7	Q8K3M7 mus musculu
40	283	15.1	343	11	Q8C131	Q8C131 mus musculu
41	283	15.1	361	13	Q90X57	Q90X57 xenopus lae
42	283	15.1	377	11	Q8BVF1	Q8BVF1 mus musculu
43	282.5	15.1	359	11	Q8BHB8	Q8BHB8 mus musculu
44	281	15.0	354	11	Q8CBJ0	Q8CBJ0 mus musculu
45	281	15.0	359	13	Q9PVY7	Q9PVY7 anguilla an

ALIGNMENTS

```
RESULT 1
Q8N5S7
ID   Q8N5S7      PRELIMINARY;      PRT;    339 AA.
AC   Q8N5S7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   G protein-coupled receptor 17.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Lung;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC031653; AAH31653.1; -.
DR   InterPro; IPR000237; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   339 AA;  37860 MW;  BB0CAFD0FC371D63 CRC64;

Query Match             20.7%;  Score 387.5;  DB 4;  Length 339;
Best Local Similarity   30.2%;  Pred. No. 1.3e-27;
Matches   94;  Conservative   57;  Mismatches  143;  Indels  17;  Gaps  5;

Qy   36 FYTALLVFSALGNIALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSS 95
         ||  :||||| :|| |||       ::|||:|:|VHLAV:||  :|  :
Db   37 FYILDFILALVGNTLALWFFIRDHKSGTPANVFLMHLAVADLSCVLVLPTRLVYHFSGNH 96

Qy   96 WPFGKGLCRLTAFVLYTDYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIW 155
         |||| | :||||:| | |: ::  |:|::          :|   | :|  :
Db   97 WPFGEIACRLTGFLFYLNMYASIYFLTCISADRFLAIVHPVKSLKLRRPLYAHLACAFLW 156

Qy  156 TLVLLQTMPLLLMPMTKPLVG3KLACME-YSSMESVLGLPLMVLVAPAIGFCGPVGIILSC 214
```

```
Db  157 VVVAVAMAPLLVSPQTVQTNHTVCLQLYREKAS-----HHALVSLAVAFTFPFITTVTC 211
            :: |    |||   | |  |||  ||:|  ||:     ||:|: :|| ||  | ::|
Qy  215 YMKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHL 274
            ||| : :|    |||||   :||  |   ||||||||| :||:|::| ||
Db  212 YLLIRSLRQGLRVE-----KRLKTKAVRMIAIVLAIFLVCFVPYHVNRSVYVLHYRSHG 266

Qy  275 PSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLGIL---KLKG--SSSS 328
            |||||||||||||||||||:||||||||||||||  ::|
Db  267 ASCATQRILALANRITSCLTSLNGALDPIMYFFVAEKPRHALCNLLCGKSRLKGPPPSFEG 326

Qy  329 SSSSSTPGKA 339
            : ||: ||:
Db  327 KTNESSLSAKS 337
```

RESULT 2
Q8VHP3
```
ID   Q8VHP3      PRELIMINARY;      PRT;    361 AA.
AC   Q8VHP3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CC-chemokine receptor 4.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jopling L.A., Sabroe I., Andrew D.P., Mitchell T.J., Li Y.,
RA   Hodge M.R., Williams T.J., Pease J.E.;
RT   "The identification, characterization and distribution of guinea pig
RT   CCR4 and epitope mapping of a blocking antibody.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF43197l; AAL57488.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE  361 AA;  41064 MW;  9304E897D4FD6839 CRC64;
```

```
Query Match             17.7%;    Score 332.5;  DB 11;  Length 361;
Best Local Similarity   27.1%;    Pred. No. 1.7e-22;
Matches   93;  Conservative   67;  Mismatches 136;  Indels 47;  Gaps  8;

Qy   15 KTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAV 74
            :|:  :       ||| :   :|||  ||  |:  ::||| |  |||||||||||||
Db   32 KTGVKAFG-----ELFLPPL-YSLVFLFGLLGNSVVVLVLFKYKRLRSMTDVYLLNLAI 84

Qy   75 SDLLFTVALP--GRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAV 132
            ||||  :||    |||||:||||||||||||| |:||||||||||||||||||||| 
Db   85 SDLLFVLSLPWG---YYAADQWVFGLGVCKMISWIYLVGFYSGIFFIVLMSIDRYLAI 140

Qy  133 VCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGL 192
            | ||  ||:    :|    ||:|:|:|  ::||| :|: |   :|  ||   ::| :||
Db  141 VHGVFSMRVRFFTYGVITSLATWAVASLPGLLFSTCYTERNHTSCKTRYSANSTTWK 200

Qy  193 PLMVIVAFAIGFCGPGVGIILSCYMKI--TWKLCSTAREDPVTSRKGRHWRGCLLTLLMLV 250
            || |:|  |:     ||| :|||| :|   |||| ||||||:|||||||||| ::| : 
Db  201 VLSSLEINILGLVIPIGIMLFCYSMIIRTLQHCKSKKKNKAVK--------MIF 246

Qy  251 AVVVCF------SPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYF 306
            |||:||      |  ::|:||||||||||||||||||||||||||: |:||||||||||||
Db  247 AVVLFLGFWTPYNIVLFLYTLVELEVLQDCSLEKYLDFALQAETLAFIHCCLNPIIYF 306

Qy  307 FASTHYRKWLLGILK------------LKGSSSSSSSSSST 335
            |    ::|    |||              :|||| :|:||
Db  307 FLGBEKFRKYIVQLFKTCRGFPAPCQYCALLRIYSNDTSSSSYT 349
```

RESULT 3
Q9BYX5
```
ID   Q9BYX5      PRELIMINARY;      PRT;    355 AA.
AC   Q9BYX5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CC chemokine receptor 8.
GN   CCR8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   DeVries M.E., Xu L., Kelvin A.A., Kelvin D.J.;
RT   "Identification and characterization of the CX3CR1 and CCR8
RT   promoters";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR   EMBL; AY016370; AAK08628.1; -.
DR   InterPro; IPR004068; CC chemkine8.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR01530; CHEMOKINER8.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Receptor; Transmembrane.
SQ   SEQUENCE  355 AA;  40830 MW;  D4167F3433B75042 CRC64;
```

```
Query Match             17.3%;    Score 324;  DB 4;  Length 355;
Best Local Similarity   26.4%;    Pred. No. 1e-21;
Matches   89;  Conservative   65;  Mismatches 145;  Indels 38;  Gaps  8;

Qy   26 QTVKIFLRSLFYTALLVFSALGN--ILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
            ||||| :||| ||||| :: :||  || :  |||||||:||||| ||||:||||||:
Db   31 QTNGKLLAVFYCLLFVFSLLGNSLVIVLVVCKKLRSI--TDVYLLNLALSDLLFVFSF 88

Qy   84 PGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVCCAHWGPRLRT 143
            |||| | ||||||  |||| ||||:||||||||||||||||| :|| :|| ::
Db   89 PFQT--YYLLDQWVFGTVMCKVVSGFYYIGFYSSMFFITLMSVDRYLAVVHAVYALKVRT 146

Qy  144 AGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVIVAFAIG 203
            ||: :  :|||:|:::
Db  147 IRMGTTLCLAVWLTAIMATIPLLLVFVFYQVASEDGVLQCYSFYNQQT---LKWKIFTNFKMN 203

Qy  204 FCG---PVGIILSCYMITWKL--CSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFSP 258
            |      ||||:|||| :||        ||||:|||||||||| ::| : :::
Db  204 ILGLLIPFTIFMFCYIKILHQLKRCQNHKT---------KAIRLVLIVIIASLLFWVP 253

Qy  259 YHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWL-- 316
            ||:|:||||||||||||||||||||||||  :| :||:   |||::|||  ::|
Db  254 FNVLFLTSLHSMHILDGCSISQQLTYATHVTEIISFTHCCVNPVIYAFVPVGEKFKKHLSE 313

Qy  317 ------------LGILKLKGSSSSSSSSSSSSTPGKASS 341
                        :|||| 
Db  314 IFQKSCSQIFNYLGRQMPRESCEKSSSSCQQHSRSRSS 350
```

RESULT 4
Q8IV19
```
ID   Q8IV19      PRELIMINARY;      PRT;    337 AA.
AC   Q8IV19;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Primary B-Cells from Tonsils;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ272207; CAC03715.1; -.
DR   EMBL; AF411112; AAL26483.1; -.
DR   EMBL; AB083600; BAB89313.1; -.
DR   EMBL; BC033571; AAH35711.1; -.
DR   Genew; HGNC:13307; GPR92.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   372 AA;  41346 MW;  BA35709084B6D84 CRC64;

Query Match             16.9%;  Score 316.5;  DB 4;  Length 372;
Best Local Similarity   30.2%;  Pred. No. 5.2e-21;
Matches  98;  Conservative  59;  Mismatches 150;  Indels  17;  Gaps   9;

Qy     32 LRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYV    91
              ||  :|| ::|| || |    ::|     : :|:  | | ||:::| |::
Db     24 LHLVVYSLVLAAGLPLNALALWVFLRALRVHSVVSVTMCNLAASDLLFTLSLPVRLSYYA    83

Qy     92 LGGSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSDHYPAVVCAHWGPRLRTAGRARLVC   151
             |  :  | |: |  |  ::      :| | ||: |  ||  || : :| :
Db     84 L-HHWPFPDLLCQTTGAIFQMNMYGSCIFMLINVDRYAAIVHPLRLRHLRRPVRARLLC   142

Qy    152 VAIWTLVLLQTMPLLLMPMT-----KPIVGKLACMEYSSMESVLG--LPLMVLVAFAIGF   204
             : |  |::|  |      :     |: |  | ||  ::::|    :  ||::|| |  ||
Db    143 LGVWALILVFAVPAARVHRPSRCRYRDLEVRL-CFESFSDELWKGRILPL-VLLABALGF   200

Qy    205 CGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIK   264
             :|| || | :: ||  |: : | ||::||    : ||  : :::: | ::: :::
Db    201 LLPLAAVVYSSGRVFWTL--ARPDATQSQRRR--KTVRLLLANIVIFLLCFVPYNSTLA   255

Qy    265 QF-MARGMLHLPSCAERRAFLLSLQATVALMMNNCGITPIIYFFASTHYRKWLLGILKLK   323
             ::  ::  :  | | || || || :   |: ||:|:| |||||||:| |||| |:|
Db    256 VYGLLRSKLVAASVPARDRVRGVLMVMVLLAGANCVLDPLVYFSAEGFRNTLRG-LGTP   314

Qy    324 GSSSSSSSSTPGKASSETPSIT   347
          |:::::::|::|:::::|:||
Db    315 HRARTSATNGTRAALAQSERSAVT   338

RESULT 6
Q91ZH4
ID   Q91ZH4;                   PRELIMINARY;       PRT;   360 AA.
AC   Q91ZH4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   C-C chemokine receptor 4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LEW;
RA   Garcia G.E., Chen S., Xia Y., Harrison J., Wilson C.B., Johnson R.J.,
RA   Bacon K.B., Feng L.;
RT   "Mononuclear cell-infiltrate inhibition by blocking macrophage-derived
RT   chemokine results in attenuation of developing crescentic
RT   glomerulonephritis";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF432872; AAL30398.2; -.
KW   Receptor.
SQ   SEQUENCE   360 AA;  41116 MW;  A71EC9D402B7FC06 CRC64;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC035750; AAH35750.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   337 AA;  38486 MW;  B9B09CE0FD305245 CRC64;

Query Match             17.2%;  Score 323;  DB 4;  Length 337;
Best Local Similarity   29.9%;  Pred. No. 1.2e-21;
Matches  94;  Conservative  50;  Mismatches 150;  Indels  .20;  Gaps   6;

Qy     34 SLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCYVLG    93
             | :::| || || |    :|:|     | :  | | ||:::|    |::|||||
Db     27 STLYSMISVVGFFGNGFVLYLIKTYHKSAFQVYMINLAVADLLCVCTLPLRVVYYVHK    86

Qy     94 SSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSDHYPAVVCAHWGPRLRTAGRARLVCVA   153
             |  |  | |: |  |: ::      :| | ||: |  ||  || : :| : :
Db     87 GIWLFGDFLCRLSTYALYYNLYCSIFFMTAMSFFRCIAIVFPVQNINLVTQKKARFVCVG   146

Qy    154 IWTLVLLQTMPLLLMPMTKPLVGKLACME---YSSMESVLGLPLMVLVAFAIGFCGPVG    209
             ||  :|||  |     :|: | | | |      ::: :|  ||:: ||| ||   :|
Db    147 IWIFVILTSSPFLMAKPQDEKNNTKCFPPQDNQTKNHVL---VLHYVSLFVGFIIPFV    203

Qy    210 IILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMAR   269
             | |:||:  |:|    :  | ::|    : ||  : : :: | :|| :: : : :
Db    204 IIIVCYTMILTLLKKSMKLSS---HKKAIGMIMVVTAAFLVSFMPYH--IQRTIHL    257

Qy    270 GMLH--LPSCAERRAFLLSLQATVALMMNNCGITPIIYFFASTHYRKWLLGILKLKGSSS   327
             |: |   || || || || ::      :|:| |||||||:| |||| |:|| :::
Db    258 HFLHNETKPCDSVLTMQKSVVITLSLAASNCCFDPLLYFFSGGNFRKGRLSTFRK-----H   312

Qy    328 SSSSSSTPGKASS   341
             | :: :|||||
Db    313 SLSSVTYVPRKKAS   326

RESULT 5
Q9H1C0
ID   Q9H1C0;                   PRELIMINARY;       PRT;   372 AA.
AC   Q9H1C0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative G protein-coupled receptor 92 (Putative G-protein coupled
DE   receptor).
GN   GRP92 OR GPR93 OR GPCR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20517346; PubMed=11062477;
RA   White K.E., Evans W.E., O'Riordan J.L.H., Speer M.C., Econs M.J.,
RA   Lorenz-Depiereux B., Grabowski M., Meitinger T., Strom T.M.,
RT   "Autosomal dominant hypophosphataemic rickets is associated with
RT   mutations in FGF23.";
RL   Nat. Genet. 26:345-348(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21458557; PubMed=11574155;
RA   Lee D.K., Nguyen T., Lynch K.R., Cheng R., Vanti W.B., Arkhitko O.,
RA   Lewis T., Evans J.F., George S.R., O'Dowd B.F.;
RT   "Discovery and mapping of ten novel G protein-coupled receptor
RT   genes.";
RL   Gene 275:83-91(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Takeda S., Kadowaki S., Haga T., Takaesu H., Mitaku S.;
RT   "Identification of G protein-coupled receptor genes from the human
RT   genome sequence.";
```

```
Query Match            16.8%;     Score 314.5; DB 11; Length 360;
Best Local Similarity  28.1%;     Pred. No. 7.7e-21;
Matches    93;   Conservative    60; Mismatches 133; Indels    45; Gaps    10;

Qy    29 KIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP--GR 86
          :||:|  : |:  :|||||||||||||||||||||||||  :|:||||||  ||
Db    39 EVFLPPL-YSLVFLLGLFGNSVVVLVLFRYKRLKSMTDVYLLNLAISDLLFVLSLPFWG- 96

Qy    87 VVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGR 146
          |||  |||||||| |||  :  ::: :||
Db    97 --YYAADQWVFGLGLCKIISWMYLVGFYSGIFFIMLSIDRYLAIVHAVFSLRARTLTY 153

Qy   147 ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFCG 206
Db   154 GVITSLITWSVAVFASLPGLLFSTCDTENNHTYCKTQYSVNSTTWKVLSSLEINVLGLVI 213

Qy   207 PVGIILSCYMKI--TWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCF----SPYH 260
          |:||:|   ||  |    ||
Db   214 PLGIMLFCYSMIIRTLRHCK------NEKKNRAVR------MIFAVVVLFLGFWTPY- 258

Qy   261 LNIKQFMAR--GMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWLLG 318
          :|:  :||
Db   259 -NVVLFLETLVELEVLQDCTLERYLDYAIQATETLAFIHCLNPVIYFFIGEKFKYIAQ 317

Qy   319 ILK----------LKGSSSSSSSSST 335
          :|:
Db   318 LFRTCRGPPVLCQHCDFLQISSADTSSSSYT 348


RESULT 7
Q8HZN4
ID   Q8HZN4;     PRELIMINARY;     PRT;    355 AA.
AC   Q8HZN4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Interleukin 8 receptor B CXCR2.
OS   Cercopithecus pygerythrus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60710;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA   Baldwin J.J., Auld D.S.;
RT   "Orthologs of human receptors and methods of use.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF540793; AAN17319.1; -.
KW   Receptor.
SQ   SEQUENCE    355 AA;    40381 MW;    D464957687BFBEBF CRC64;

Query Match            16.7%;     Score 313.5; DB 6; Length 355;
Best Local Similarity  28.5%;     Pred. No. 9.4e-21;
Matches    95;   Conservative    58; Mismatches 153; Indels    27; Gaps    8;

Qy    24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
          |:| :|:|||:  : |:  :||||||||
Db    37 ESLEINKYFVIIYALVFLLSLLGNSLVMLVILHSRVGRSVTDVYLLNLAMADLLFALTL 96

Qy    84 PGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT 143
          | :    |   ::: :|| ||
Db    97 P--IWAAAKNGWIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHA---TRTLT 151

Qy   144 AGR--ARLVCVAIWTLVLLQTMPLLLMPMT-KPLVGKLACMEYSSMESVLGLPLMVLVAF 200
          |:|
Db   152 QKRYLVKFVCLSIWGLSLLLALPVLLFRRTVYPTYISPVCYEDMGNNTAKWRMVLRILPQ 211

Qy   201 AIGFCGPVGIILSCY---MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS 257
          :|: :||
Db   212 TFGFIPLLIMLFCYGFTLRTLFK---------AHMGQKHRAMRVIFAVVLIFLLCWL 260

Qy   258 PYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWL 316
```

```
Db   261 PYHLVLTDTLMRTLRIKETCQRRNDIRDALDATEILGILHSCLNPIIYAFIGQKFRHGL 320
          :||||   :  :    ::  :        ||||    :||||
Qy   317 LGILKLKGSSSSSS---SSSTPGKASSETPS 345
Db   321 LKILATHGLISKDSLPKDSRPSFVGSSGHTSS 353


RESULT 8
Q8HZN5
ID   Q8HZN5;     PRELIMINARY;     PRT;    355 AA.
AC   Q8HZN5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Interleukin 8 receptor B CXCR2.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA   Baldwin J.J., Auld D.S.;
RT   "Orthologs of human receptors and methods of use.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF540792; AAN17318.1; -.
KW   Receptor.
SQ   SEQUENCE    355 AA;    40293 MW;    6A9BF9A921F470C5 CRC64;

Query Match            16.6%;     Score 310.5; DB 6; Length 355;
Best Local Similarity  28.7%;     Pred. No. 1.8e-20;
Matches    95;   Conservative    58; Mismatches 151; Indels    27; Gaps    8;

Qy    24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
          |:| :|:|||:  : |:  :||||||||
Db    37 ESLEINKYFVIIYALVFLLSLLGNSLVMLVILHSRVGRSITDVYLLNLAMADLLFALTL 96

Qy    84 PGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT 143
          |      |    ::: :|| ||
Db    97 PIWAAAKVNG--WIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHA---TRTLT 151

Qy   144 AGR--ARLVCVAIWTLVLLQTMPLLLMPMTKPLVG-KLACMEYSSMESVLGLPLMVLVAF 200
          |:|
Db   152 QKRYLVKFVCLSIWSLSLLLALPVLLFRRTVYLTYISPVCYEDMGNNTAKWRMVLRILPQ 211

Qy   201 AIGFCGPVGIILSCY---MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAVVVCFS 257
          :|: :||
Db   212 TFGFIPLLIMLFCYGLTLRTLFK---------AHMGQKHRAMRVIFAVVLIFLLCWL 260

Qy   258 PYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKWL 316
          :||||   :  :    ::  :        ||||    :||||
Db   261 PYHLVLLADTLMRTLRLINETCQRRNNIDQALDATEILGILHSCLNPLIYAFIGQKFRHGL 320
Qy   317 LGILKLKGSSSSSS---SSSSTPGKASSET 343
Db   321 LKILATHGLISKDSLPKDSRPSFVGSSSGHT 351


RESULT 9
Q8HZN6
ID   Q8HZN6;     PRELIMINARY;     PRT;    355 AA.
AC   Q8HZN6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Interleukin 8 receptor B CXCR2.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA   Baldwin J.J., Auld D.S.;
RT   "Orthologs of human receptors and methods of use.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF540791; AAN17317.1; -.
KW   Receptor.
SQ   SEQUENCE   355 AA;   40245 MW;   448C5848EB559AC8 CRC64;

     Query Match    16.5%;   Score 309.5;   DB 6;   Length 355;
     Best Local Similarity   28.0%;   Pred. No. 2.2e-20;
     Matches   93;   Conservative   64;   Mismatches  146;   Indels  29;   Gaps   9;

Qy   24  ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL 83
             : :|| ||| ||||  | |: : |:  : | ||||||:|||::||||||
Db   37  ESLEINKYFVVIYLVLVPFLLSLLGNSLVMLVILYSRVGRSVTDVYLLNLALADLLFALTL 96

Qy   84  PGRVVCVYLGSSWNPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRT 143
             |  |  |  ||| | |: ||| | |||||||||||||||||||||||||||||
Db   97  PIWAASKVTG--WIFGTFLCKVSLLKEVNFYSGILLLACISVDRYLAIVHA---TRTLT 151

Qy  144  AGR--ARLVCVAIWTLVLLQTMPLLLMPMT--KPLVGKLACMEYSSMESVLGLPLMVLVA 199
             :  :|: :|:  || | :  :: |||:  |:|| | | :  |::|: ::
Db  152  QKRYLVKFICLSIWGLSLLLALPVLIFRKTIYPPYVSPV-CYEDMGNNTANWRMLLRILP 210

Qy  200  FAIGFCGPVGIILSCY--MKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCF 256
             : :||||||| ||      ||: |||||||||||||||||||| : :: :
Db  211  QSFGFIVPLLIMLFCYGFTLRTLFK----------AHMGQKHRAMRVIFAVVLIFLLCW 259

Qy  257  SPYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYRKW 315
             |:|    |  || |    : ::   |  :    : ::|| :||||||||||:| |
Db  260  LPTNLVLLADTLMRTWVIQETCERRNDIDRALEATEILGILHSCLNPLIYAFIGQKFRHG 319

Qy  316  LLGILKLKGSSSSSS-----SSSSTPGKASSET 343
             ||: ||||:  |        : :| |  ::|
Db  320  LLKILAIHGLISKDSLPKDSRPSFVGSSSGHT 351


RESULT 10
Q9MYJ8
ID   Q9MYJ8;      PRELIMINARY;      PRT;      355 AA.
AC   Q9MYJ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Chemokine receptor.
GN   CCR1.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20153429; PubMed=10686294;
RA   Liang M., Rosser M.P., Ng H.P., May K., Bauman J.G., Islam I.,
RA   Ghannam A., Kretschmer P.J., Pu H., Dunning L., Snider R.M.,
RA   Morrissey M.M., Hesselgesser J., Perez H.D., Horuk R.;
RT   "Species selectivity of a small molecule antagonist for the CCR1
RT   chemokine receptor.";
RL   Eur. J. Pharmacol. 389:41-49(2000).
DR   EMBL; AF127528; AAF36453.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   355 AA;   40928 MW;   2B01C47E9874A2C1 CRC64;

     Query Match    16.5%;   Score 309;   DB 6;   Length 355;
     Best Local Similarity   27.0%;   Pred. No. 2.5e-20;
     Matches   87;   Conservative   61;   Mismatches  132;   Indels  42;   Gaps   7;
```

```
Qy   37  YTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALPGRVVCVYLGSSW 96
             |: || |||| :|||||||||||||||:||||||||:|||||||| |  |  |  |||
Db   41  YSLVFVIGLVGNILVVVVLVQYKRLKNMTSIYLLNLAISDLLFLFTLP-FWISYQLKTDW 99

Qy   97  PFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWT 156
             ||   | | : :| | ||||||||:|| |||: :: |::  :| :| |   :|||:|
Db  100  VFGNAMCKVLSGFYYTGLYSEIFFILLLTIDRYLAIVHAVFALRARTVTFGVITSIIIWV 159

Qy  157  LVLLQTMPLLLMPMTKPLVGKLACMEYSSMES---------VLGLPLMVLVAFAIGFCGPV 208
             |::| : ||| : : : : :: |                  |:|| | : :|   :|
Db  160  LAILASLPGLYFAKTQWEITHRTCSLHFPHESRQEWKLFQALKLNLLGLVL------PL 212

Qy  209  GIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMA 268
             |: :|||| |   |||||| : :   ::||:       |: :: ::  ||||: : ::
Db  213  LVMIVCYTGIKILKILLRRPNE----KKSKAVR---LIFVIMIIFLFWTPYNLTTLISVF 264

Qy  269  RGMLHLPSCAERRAFLLSLQATVALMNMNCGITPIIYFFASTHYR------------- 313
             : : |  :   |    :  ||: | :: |||| ::::||  ||
Db  265  QDFLFTYGCEGRQLDLAIQVTEMIAYTHCCVNPVIYAFVGERFRKHLRQLFHRRVAVHL 324

Qy  314  -KWL--LGILKLKGSSSSSSSS 332
              |:|  || ||||:::::::::
Db  325  VKWLPFLSVDRLERASSTSPST 346


RESULT 11
Q8MJW8
ID   Q8MJW8;      PRELIMINARY;      PRT;      360 AA.
AC   Q8MJW8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CC chemokine receptor 4.
GN   CCR4.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Maeda S., Okayama T., Masuda K., Ohno K., Tsujimoto H.;
RT   "Detection of CC chemokine receptor 4 (CCR4) mRNA expression in canine
RT   atopic skin lesion.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB080188; BAC10546.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
SQ   SEQUENCE   360 AA;   41354 MW;   69115F5209EC0908 CRC64;

     Query Match    16.4%;   Score 308.5;   DB 6;   Length 360;
     Best Local Similarity   27.8%;   Pred. No. 2.8e-20;
     Matches   92;   Conservative   59;   Mismatches 135;   Indels  45;   Gaps   9;

Qy   29  KIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVALP--GR 86
Db   39  ELFLPPL-YSLVFLFGLLGNSVVVVVLFKYKRLKSMTDVLLNLAISDLLFVLSLPFWG- 96

Qy   87  VVCVYLGSSWNPFGKGLCRLTAFVLYTDTYGGVYLMACVSVDHYPAVVCAHWGPRLRTAGR 146
Db   97  ---YYAADQWVFGLGLCKIISWMYLVGFYSGIFFIMLMSIDRYLAIVHAVFSLRARTLTY 153

Qy  147  ARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAFAIGFPCG 206
Db  154  GVITSLATWSVAVLASLPGLLFSTCYTERNHTYCKTKYSRNSTRMKVLSSLEINILGLVI 213

Qy  207  PVGIILSCYMKI--TWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVCF----SPYH 260
Db  214  PLGTMLFCYSMIIRTLQHCKNEKKSK-------------AVRMVFAVVALFLGPWAPY- 258
```

```
RA   Horlick R.A., Zhao J., Swanson R.N., Webb M.L., Strohl B.,
RA   Baldwin J.J., Auld D.S.;
RT   "Orthologs of human receptors and methods of use.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF540789; AAN17315.1; -.
KW   Receptor.
SQ   SEQUENCE   355 AA;   40122 MW;   809AB63ADA6A42C6 CRC64;

Query Match             16.3%;  Score 306.5;  DB 6;  Length 355;
Best Local Similarity   28.0%;  Pred. No. 4.2e-20;
Matches   94;  Conservative   61;  Mismatches  144;  Indels   37;  Gaps    9;

Qy     24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL  83
                 :  ::|||.|:|||    :.|:|:||||:.|||||
Db     37 ESLEINKYFVVIYALVFLLSLLGNSLVMLVILYSRVGRSVTDVYLLNLALADLLFALTL  96

Qy     84 PGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSDHYPAVVCAHWGPRLRT 143
                 : :    :  : |:|    .: :.  ::|:||
Db     97 PIWAASKVNG--WIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHA--TRTLT 151

Qy    144 AGR--ARLVCVAIWTLVLLQTMPLLLMPMT-----KPLVGKLACMEYSSMESVLGLPLM 195
                   :.:|:|   :|:|||      :||||
Db    152 QKRYLVKFICLSIWGLSLLLALPVLLFRRTVYSSNVSP-----ACYEDMGNNTANWRMLL 206

Qy    196 VLVAFAIGFCGPVGIILSCY---MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAV 252
                 :||:||||       ::.|     :|:
Db    207 RILPQSFGFIVPLLIMLFCYGFTLRTLFK----------AHMGQKHRAMRVIFAVVLIF 255

Qy    253 VVCFSPYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTH 311
                 :|:||||    |||    ::|:|:|||||||
Db    256 LLCWLPYNLVLLADTLMRTQVIQETCERRNHIDRALDATEILGILHSCLNPLIYAFIGQK 315

Qy    312 YRKWLLGILKLKGSSSSSS---SSSSTPGKASSET 343
                 | |:|   ||
Db    316 FRHGLLKILAIHGLISKDSLPKDSRPSFVGSSSGHT 351


RESULT 14
Q8NE10
ID   Q8NE10                    PRELIMINARY;       PRT;    362 AA.
AC   Q8NE10;
DT   01-OCT-2002  (TrEMBLrel. 22, Created)
DT   01-OCT-2002  (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003  (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC036661; AAH36661.1; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   362 AA;   41478 MW;   B0C0EC1AE3F81FB8 CRC64;

Query Match             16.3%;  Score 306.5;  DB 4;  Length 362;
Best Local Similarity   23.9%;  Pred. No. 4.3e-20;
Matches   79;  Conservative   75;  Mismatches  148;  Indels   29;  Gaps    6;

Qy     22 NVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTV  81
                 |  ||  :  || :|||| |    :||.||:|||||
Db     36 NMPNKSVLLYTLSFIYIFIVIGMIANSVVWNIQAKTTGYDTHCYILNLAIADLWVVL  95

Qy     82 ALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSDHYPAVVCAHWGPRL 141
                 | ||    |||    : || :  :
```

```
Qy    261 LNIKQFMAR--GMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRKWLLG 318
                 .:    :  :|  :|:|:||||||
Db    259 -NVVLFLETLVELVLQDCTFERHLDYAIQATETLAFVHCCLNPVIYFFLGEKFRKYLVQ 317

Qy    319 ILK----------LKGSSSSSSSSSST 335
                 |||
Db    318 LFKCRGPPMLCQYCRLLQMYSPDTPSSYT 348


RESULT 12
Q8IUZ1
ID   Q8IUZ1                    PRELIMINARY;       PRT;    355 AA.
AC   Q8IUZ1;
DT   01-MAR-2003  (TrEMBLrel. 23, Created)
DT   01-MAR-2003  (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003  (TrEMBLrel. 23, Last annotation update)
DE   Interleukin 8 receptor, beta.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC037961; AAH37961.1; -.
KW   Receptor.
SQ   SEQUENCE   355 AA;   40122 MW;   809AB63ADA6A42C6 CRC64;

Query Match             16.3%;  Score 306.5;  DB 4;  Length 355;
Best Local Similarity   28.0%;  Pred. No. 4.2e-20;
Matches   94;  Conservative   61;  Mismatches  144;  Indels   37;  Gaps    9;

Qy     24 ESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSRKINCTGIYLVHLAVSDLLFTVAL  83
                 :  ::|||.|:|||    :.|:|:||||:.|||||
Db     37 ESLEINKYFVVIYALVFLLSLLGNSLVMLVILYSRVGRSVTDVYLLNLALADLLFALTL  96

Qy     84 PGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYLMACVSDHYPAVVCAHWGPRLRT 143
                 : :    :  : |:|    .: :.  ::|:||
Db     97 PIWAASKVNG--WIFGTFLCKVVSLLKEVNFYSGILLLACISVDRYLAIVHA--TRTLT 151

Qy    144 AGR--ARLVCVAIWTLVLLQTMPLLLMPMT-----KPLVGKLACMEYSSMESVLGLPLM 195
                   :.:|:|   :|:|||      :||||
Db    152 QKRYLVKFICLSIWGLSLLLALPVLLFRRTVYSSNVSP-----ACYEDMGNNTANWRMLL 206

Qy    196 VLVAFAIGFCGPVGIILSCY---MKITWKLCSTAREDPVTSRKGRHWRGCLLTLLMLVAV 252
                 :||:||||       ::.|     :|:
Db    207 RILPQSFGFIVPLLIMLFCYGFTLRTLFK----------AHMGQKHRAMRVIFAVVLIF 255

Qy    253 VVCFSPYHL-NIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTH 311
                 :|:||||    |||    ::|:|:|||||||
Db    256 LLCWLPYNLVLLADTLMRTQVIQETCERRNHIDRALDATEILGILHSCLNPLIYAFIGQK 315

Qy    312 YRKWLLGILKLKGSSSSSS---SSSSTPGKASSET 343
                 | |:|   ||
Db    316 FRHGLLKILAIHGLISKDSLPKDSRPSFVGSSSGHT 351


RESULT 13
Q8HZN8
ID   Q8HZN8                    PRELIMINARY;       PRT;    355 AA.
AC   Q8HZN8;
DT   01-MAR-2003  (TrEMBLrel. 23, Created)
DT   01-MAR-2003  (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003  (TrEMBLrel. 23, Last annotation update)
DE   Interleukin 8 receptor B CXCR2.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
```

Mon Feb 23 11:38:30 2004   us-10-041-615-34.rspt   Job time : 43 secs

```
Db     96 TIPWVVSLVQHNQWPMGBLTCKVTHLIFSINLFGSIPFLTCMSVDRYLSITYFTNTPSS 155

Qy    142 RTAGRARLVCVAIWTLVLLQTMP--LLLMPMTKPLVGKLACM----EYSSMESVLGLPLM 195
           ::|||: :||       ::|       :|    ::|| ||:|||
Db    156 RKKMVRRVVCILWLLAFCVSLPDTYYLKTVTSASNNETYCRSFYPEHSIKEWLIG---M 212

Qy    196 VLVAPAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWRGCLLTLMLVAVVVC 255
           ||: :||  |        :|                      :|   ||| : :||
Db    213 ELVSVVLGFAVFFSIVAVFYFLLARAISASSDQEKHSSRK------IIFSYVVVFLVC 264

Qy    256 FSPYHLNIKQFMARGMLHLP-SCAERRAFLLSLQATVALMNNCGITPIIYFFASTHYRK 314
           | ||||: : :||:  :    :|||||||||||||||:| :|::|::|||:|:| ||
Db    265 WLPTHVAVLLDIFSILHYIPFTCRLEHALFTALHVTQCLSLVHCCVNPVLYSFINRNYRY 324

Qy    315 WLJ---------GILLLKGSSSSSSSSS 334
           |:         |:|| |
Db    325 ELMKAFIFKYSAKTGLTKLIDASRVSETEYS 355


RESULT 15
Q8BY68
ID   Q8BY68          PRELIMINARY;          PRT;     352 AA.
AC   Q8BY68;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to putative purinergic receptor P2Y10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK041740; BAC31049.1; -.
SQ   SEQUENCE   352 AA;  40382 MW;  C9977E40202AB4EA CRC64;

Query Match          16.3%;    Score 305;   DB 11;   Length 352;
Best Local Similarity   27.9%;   Pred. No. 5.7e-20;
Matches   84;   Conservative   59;   Mismatches  130;   Indels   28;   Gaps   9;

Qy     27 TVKIFLRSLF---YTALLVFSALGNILALCLTCQXSRKINCTGIYLVHLAVSDLLFTVAL 83
           | ||| :|         :|| :| |: |   ||:|  ::|||:|:::|||| :|
Db     16 TYLTPQYSLYATTYIFIFIPGLLANSVALWVLCRFISKKNKAIIFMINLSVADLAHVLSL 75

Qy     84 PGRVVCVLGSSWPFGKGLCRLITAFVLYTDTYGGVYLMACVSVDHYPAVV----CAHWGP 139
           | |: ||| ||||| ||| :||:::  |||:  : ::|| :
Db     76 PLRIY-YYINRHWPFQRALCLLCFVLKYLNMYASIFFLTCISLQRCLFLLKPFRARNWKR 134

Qy    140 RLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC-----MEYSSMESVLGLPL 194
           | |:          :    ::|  :   |:  :||||||||         ::|||:|:
Db    135 RYDVA----ISAAVWIVGTACLPLPILRSAGLANNSESCFADLGLQDITMASSIG--- 186

Qy    195 MVLVAFAIGFCGPVGIILSCYMKITWKLCSTARE--DPVTSRKGRHWRGCLLTLMLVAV 252
           |:|| |||| | ||:  :                 :||||||||| || |||||| :|
Db    187 MVTAAELGGFVLPVVIITYC----TWKTRKSLQEFQDPPQNIKERK-KALRMVLMCAVVF 241

Qy    253 VVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNCGITPIIYFFASTHY 312
           :||::|||||||||        :|||||||||||||||||:| :|::|::|||:|:|
Db    242 IVCFTPYHLNFPFFMVKQRVFSNCSFIRNTLCFHIISLCLANLNCCLDPVVYFMTSEF 301

Qy    313 R 313
           R
Db    302 R 302
```

Search completed: February 19, 2004, 19:34:23

```
             GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 19, 2004, 19:33:35 ; Search time 43 Seconds
                                     (without alignments)
                             1328.874 Million cell updates/sec

Title:        US-10-041-615-34
Perfect score: 360
Sequence:   1 MIKLGPYFTPPTKIKTKIKD........SETPSITQARGSMFLAEHVV 360

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched:     1107863 seqs, 158726573 residues

Word size :    10

Total number of hits satisfying chosen parameters:     692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :    A_Geneseq_19Jun03:*
   1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
   2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
   3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
   4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
   5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
   6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
   7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
   8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
   9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
  10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
  11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
  12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
  13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
  14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
  15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
  16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
  17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
  18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
  19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
  20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
  21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
  22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
  23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
  24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	360	100.0	360	23	ABG32286	Human GPCR16 prote
2	86	23.9	110	23	ABG66698	Human novel polype
3	86	23.9	377	22	ABG14682	Novel human diagno
4	70	19.4	177	22	AAU19222	Human G protein-co
5	51	14.2	177	22	AAU25588	Human G Protein-Co
6	13	3.6	311	22	ABB67853	Drosophila melanog
7	12	3.3	106	22	AAU32649	Novel human secret
8	12	3.3	1262	22	ABB70839	Drosophila melanog
9	11	3.1	85	22	AAU32705	Novel human secret
10	11	3.1	86	23	ABP10952	Human ORFX protein
11	11	3.1	94	23	AAU16320	Human novel secret
12	11	3.1	94	24	ABU55389	Human novel polype
13	11	3.1	104	22	ABG58455	Human liver peptid
14	11	3.1	104	22	ABB43053	Peptide #10559 enc
15	11	3.1	104	22	ABB26224	Protein #8223 enco
16	11	3.1	104	22	AAM63955	Human brain expres
17	11	3.1	104	22	AAM76775	Human bone marrow
18	11	3.1	104	22	AAM20996	Peptide #7430 enco
19	11	3.1	104	22	AAM36882	Peptide #10919 enc
20	11	3.1	104	23	ABG45940	Human peptide enco
21	11	3.1	112	22	AAU15859	Human novel secret
22	11	3.1	112	22	ABU54928	Human novel polype
23	11	3.1	115	22	AAU32633	Novel human secret
24	11	3.1	115	22	AAO00547	Human polypeptide
25	11	3.1	115	22	AAO00737	Human polypeptide
26	11	3.1	118	22	AAU32634	Novel human secret
27	11	3.1	134	22	AAG91288	C glutamicum prote
28	11	3.1	141	22	ABB69191	Drosophila melanog
29	11	3.1	142	22	AAU19607	Human diagnostic a
30	11	3.1	142	23	ABG60206	Human DITHP polype
31	11	3.1	145	22	ABG18589	Novel human diagno
32	11	3.1	145	22	ABG18766	Novel human diagno
33	11	3.1	145	22	ABG19509	Novel human diagno
34	11	3.1	147	21	AAG08169	Arabidopsis thalia
35	11	3.1	148	21	AAG43674	Arabidopsis thalia
36	11	3.1	178	21	AAG08048	Arabidopsis thalia
37	11	3.1	178	21	AAG51989	Arabidopsis thalia
38	11	3.1	186	21	AAG08168	Arabidopsis thalia
39	11	3.1	187	21	AAG43673	Arabidopsis thalia
40	11	3.1	187	22	AAE05009	Cryptosporidium pa
41	11	3.1	198	22	AAO02467	Human polypeptide
42	11	3.1	199	21	AAU16338	Human novel secret
43	11	3.1	199	24	ABU55407	Human novel polype
44	11	3.1	201	22	AAU15883	Human novel secret
45	11	3.1	201	24	ABU54952	Human novel polype

ALIGNMENTS

```
RESULT 1
ABG32286
ID   ABG32286 standard; Protein; 360 AA.
XX
AC   ABG32286;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Human GPCR16 protein.
XX
KW   G protein coupled receptor related protein; human; GPCR;
KW   cardiomyopathy; atherosclerosis; diabetes; cancer; stroke;
KW   von Hippel-Lindau syndrome; Alzheimer's disease; tuberous sclerosis;
KW   hypercalcaemia; Parkinson's disease; Huntington's disease;
KW   cerebral palsy; epilepsy; Lesch-Nyhan syndrome; multiple sclerosis;
KW   ataxia-telangiectasia; leukodystrophy; addiction; anxiety; depression;
KW   pain; obesity; Crohn's disease; osteoporosis; haemophilia; asthma;
KW   inflammatory bowel disease; infertility; hypertension; scleroderma;
KW   arthritis; human immunodeficiency virus; autoimmune disease; HIV;
KW   infection; graft-versus-host disease.
XX
OS   Homo sapiens.
XX
PN   WO200264793-A2.
XX
PD   22-AUG-2002.
XX
PF   03-JAN-2002; 2002WO-US00056.
XX
PR   03-JAN-2001; 2001US-259552P.
PR   09-JAN-2001; 2001US-260544P.
PR   20-MAR-2001; 2001US-277405P.
```

```
Qy      301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360

Db      301 TPIIYFFASTHYRKWLLGILKLKGSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV 360


RESULT 2
ABG6698
ID   ABG6698 standard; Protein; 110 AA.
XX
AC   ABG6698;
XX
DT   30-AUG-2002  (first entry)
XX
DE   Human novel polypeptide #33.
XX
KW   Human; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombosis; thrombolysis; .coagulation disorder;
KW   fungal infection.
XX
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US47004.
XX
PR   30-NOV-2000; 2000US-0028952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Wang J,  Wang D;
PI   Yamazaki V,  Ujwal ML,  Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   N-PSDB; ABK94922.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders; cancer and promoting wound healing -
XX
PS   Claim 10; Page 600; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABG6666-ABG6758 represent human
CC   novel polypeptides of the invention.
XX
```

```
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Casman SJ,  Edinger SR,  Ellerman K,  Smithson G,  Kekuda R;
PI   Muralidhara P;
XX
DR   WPI; 2002-643487/69.
DR   N-PSDB; ABS64723.
XX
PT   New isolated G protein coupled receptor polypeptides and
PT   polynucleotides, useful in gene therapy, particularly for treating or
PT   preventing cardiomyopathy, atherosclerosis, diabetes, multiple
PT   sclerosis, Crohn's disease or cancer -
XX
PS   Claim 1; Page 98; 252pp; English.
XX
CC   This invention relates to a new isolated G-protein coupled receptor
CC   (GPCRX) polypeptide sequence and the cDNA encoding it. The GPCRX
CC   polypeptide, GPCRX nucleic acid and an antibody specific to the protein
CC   are useful for treating, preventing or alleviating a GPCRX-associated
CC   disorder or a pathological state in a subject, particularly a human. In
CC   particular, the disorder is cardiomyopathy, atherosclerosis, diabetes,
CC   or a disorder related to cell signal processing and metabolic pathway
CC   modulation. The GPCRX polypeptide and nucleic acid are also useful for
CC   diagnosing the presence of or predisposition to a disease associated
CC   with altered levels of GPCRX, particularly cancer. The GPCRX nucleic
CC   acid and polypeptide are especially useful in the manufacture of a
CC   medicament for therapeutic or prophylactic applications for disorders
CC   associated with aberrant GPCRX expression or activity, e.g. Von
CC   Hippel-Lindau syndrome, Alzheimer's disease, stroke, tuberous
CC   sclerosis, hypercalcaemia, Parkinson's disease, Huntington's disease,
CC   cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis,
CC   ataxia-telangiectasia, leukodystrophies, addiction, anxiety, depression,
CC   pain, obesity, Crohn's disease, osteoporosis, inflammatory bowel
CC   disease, infertility, hypertension, scleroderma, haemophilia,
CC   asthma, arthritis, human immunodeficiency virus; autoimmune disease;
CC   HIV, viral, fungal, bacterial or protozoal infections, or graft-versus
CC   -host disease. The DNA encoding the protein is useful in gene therapy
CC   for treating the above conditions. The polypeptides can be used as
CC   immunogens to produce antibodies and as vaccines. The nucleic acids are
CC   further used as hybridisation probes, in chromosome mapping, tissue
CC   typing, preventive medicine, and pharmacogenomics. These are also useful
CC   in developing powerful assay system for functional analysis of various
CC   human disorders, as well as in diagnostic applications. The present
CC   sequence represents a human G protein coupled receptor related protein
CC   (GPCR) of the invention.
XX
SQ   Sequence   360 AA;
```

```
Query Match             100.0%;  Score 360;  DB 23;  Length 360;
Best Local Similarity   100.0%;
Matches  360;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MIKLGPYFTPPTKIKTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIILALCLTCQKSR 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIKLGPYFTPPTKIKTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNIILALCLTCQKSR 60

Qy     61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120

Qy    121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC 180

Qy    181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR 240

Qy    241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNNCGI 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCLLTLLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVALMNNNCGI 300
```

```
SQ   Sequence   110 AA;

Query Match          23.9%;   Score 86;  DB 23;  Length 110;
Best Local Similarity  100.0%;  Pred. No. 2.8e-75;
Matches  86;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  141 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAF 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAF 80

Qy  201 AIGFCGPVGIILSCYMKITWKLCSTA 226
            ||||||||||||||||||||||||||
Db   81 AIGFCGPVGIILSCYMKITWKLCSTA 106


RESULT 3
ABG14682
ID  ABG14682 standard; Protein; 377 AA.
XX
AC  ABG14682;
XX
DT  18-FEB-2002  (first entry)
XX
DE  Novel human diagnostic protein #14673.
XX
KW  Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200175067-A2.
XX
PD  11-OCT-2001.
XX
PF  30-MAR-2001; 2001WO-US08631.
XX
PR  31-MAR-2000; 2000US-0540217.
PR  23-AUG-2000; 2000US-0649167.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Drmanac RT, Liu C, Tang YT;
XX
DR  WPI; 2001-639362/73.
DR  N-PSDB; AAS78869.
XX
PS  Claim 20; SEQ ID No 45041; 103pp; English.
XX
CC  The invention relates to isolated polynucleotide (I) and
CC  polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC  polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC  and gene mapping, and in recombinant production of (II). The
CC  polypeptide and polynucleotide sequences have applications in
CC  diagnostics, forensics, gene mapping, identification of mutations
CC  responsible for genetic disorders or other traits to assess biodiversity
CC  and to produce other types of data and products dependent on DNA and
CC  amino acid sequences. ABG00010-ABG30377 represent novel human
CC  diagnostic amino acid sequences of the invention.
CC  Note: The sequence data for this patent did not appear in the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   377 AA;

Query Match          23.9%;   Score 86;  DB 22;  Length 377;
Best Local Similarity  100.0%;  Pred. No. 8.8e-75;
Matches  86;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  141 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAF 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLGLPLMVLVAF 175

Qy  201 AIGFCGPVGIILSCYMKITWKLCSTA 226
            ||||||||||||||||||||||||||
Db  176 AIGFCGPVGIILSCYMKITWKLCSTA 201


RESULT 4
AAU19222
ID  AAU19222 standard; Protein; 177 AA.
XX
AC  AAU19222;
XX
DT  04-DEC-2001  (first entry)
XX
DE  Human G protein-coupled receptor nGPCR-2343.
XX
KW  Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic;
KW  cytostatic; cardiant; antidiabetic; anoretic; hypotensive; hypertensive;
KW  antiparkinsonian; nootropic; neuroprotective; antidepressant;
KW  viral infection; HIV-1; human immunodeficiency virus; HIV-2; pain;
KW  cancer; metabolic disease; cardiovascular disease; type 2 diabetes;
KW  obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW  atherosclerosis; Parkinson's disease; psychosis; neurological disorder;
KW  schizophrenia; migraine; major depression; anxiety; mental disorder;
KW  manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.
XX
OS  Homo sapiens.
XX
PN  WO200166750-A2.
XX
PD  13-SEP-2001.
XX
PF  08-MAR-2001; 2001WO-US07322.
XX
PR  08-MAR-2000; 2000US-0187581.
PR  08-MAR-2000; 2000US-0187582.
PR  08-MAR-2000; 2000US-0187714.
PR  08-MAR-2000; 2000US-0187715.
PR  08-MAR-2000; 2000US-0187825.
PR  08-MAR-2000; 2000US-0187828.
PR  08-MAR-2000; 2000US-0187829.
PR  08-MAR-2000; 2000US-0187830.
PR  08-MAR-2000; 2000US-0187833.
PR  08-MAR-2000; 2000US-0187874.
PR  08-MAR-2000; 2000US-0187930.
PR  08-MAR-2000; 2000US-0188049.
PR  08-MAR-2000; 2000US-0189294.
PR  08-MAR-2000; 2000US-0187929.
PR  08-MAR-2000; 2000US-0187928.
XX
PA  (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI  Vogeli G, Wood LS;
XX
DR  WPI; 2001-536778/59.
DR  N-PSDB; AAS30791.
XX
PT  Isolated nucleic acid molecules encoding G protein-coupled receptors
PT  termed nGPCR-x, useful in the treatment and diagnosis of viral
PT  infections, cancers and mental disorders (e.g. Parkinson's disease and
PT  schizophrenia) -
XX
```

```
PS   Claim 31; Page 281; 336pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules encoding
CC   G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides,
CC   polypeptides, and modulators may be used in the treatment of diseases and
CC   conditions such as infections, such as viral infections caused by HIV-1
CC   (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and
CC   cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC   anorexia, hypotension, hypertension, myocardial infarction,
CC   atherosclerosis), Parkinson's disease, and psychotic and
CC   neurological disorders, including schizophrenia, migraine, major
CC   depression, anxiety, mental disorder, manic depression, and
CC   dyskinesias, such as Huntington's disease or Tourette's Syndrome
CC   and many other diseases and syndromes listed in the specification.
CC   nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x
CC   modulators, may also be used in diagnostic assays for such diseases or
CC   conditions. The present sequence represents a G protein-coupled
CC   receptor of the invention.
XX
SQ   Sequence   177 AA;

Query Match             19.4%;  Score 70;  DB 22;  Length 177;
Best Local Similarity   100.0%;  Pred.No.1.5e-59;
Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          233 SRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVA 292
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 SRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVA 60

Qy          293 LMNNNCGITP 302
                ||||||||||
Db           61 LMNNNCGITP 70

RESULT 5
AAU25588
ID   AAU25588 standard; Protein; 177 AA.
XX
AC   AAU25588;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human G Protein-Coupled Receptor (GPCR) polypeptide #35.
XX
KW   Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW   attention deficit disorder; anxiety; depression; bipolar disorder;
KW   neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW   metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW   type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW   cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW   viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
KW   antidepressant; anorectic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200162797-A2.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-US05676.
XX
PR   23-FEB-2000; 2000US-0184247.
PR   23-FEB-2000; 2000US-0184303.
PR   23-FEB-2000; 2000US-0184304.
PR   23-FEB-2000; 2000US-0184305.
PR   23-FEB-2000; 2000US-0184397.
PR   02-MAR-2000; 2000US-0186457.
PR   03-MAR-2000; 2000US-0186810.
PR   09-MAR-2000; 2000US-0188064.
PR   13-MAR-2000; 2000US-0188880.
PR   03-APR-2000; 2000US-0194344.
PR   23-JUN-2000; 2000US-0213861.
PR   11-JUL-2000; 2000US-0217369.

PR   11-JUL-2000; 2000US-0217370.
PR   14-JUL-2000; 2000US-0218337.
PR   20-JUL-2000; 2000US-0218492.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G, Wood LS, Parodi LA, Lind P;
XX
DR   WPI; 2001-570628/64.
DR   N-PSDB; AAS42840.
XX
PT   New isolated nucleic acid encoding a new G-protein coupled receptor.
PT   polypeptide for detecting receptor modulators that can treat mental
PT   disorders, such as schizophrenia, anxiety, depression, or obesity -
XX
PS   Claim 35; Page 83; 279pp; English.
XX
CC   Sequences AAU25554-AAU25616 represent human G-protein coupled receptor.
CC   (GPCR) polypeptides of the invention. The proteins and their associated
CC   DNA sequences can be used to identify compounds which bind to GPCR
CC   polypeptides and in screening for compounds that modulate GPCR activity.
CC   By screening a human subject for the presence of mutations in GPCR DNA, a
CC   GPCR-related disorder or a genetic predisposition can be diagnosed. The
CC   sequences can also be used for treatment and prevention of mental
CC   disorders such as schizophrenia, attention deficit disorder, anxiety,
CC   depression, dementia, bipolar disorder, neurological disorders such as
CC   Huntington's disease, Parkinson's disease and Tourette's syndrome,
CC   metabolic disorders such as obesity, anorexia and type 2 diabetes,
CC   cardiovascular disorders such as thrombosis, myocardial infarction,
CC   cardiomyopathy and atherosclerosis, viral infections caused by HIV and
CC   cancers.
XX
SQ   Sequence   177 AA;

Query Match             14.2%;  Score 51;  DB 22;  Length 177;
Best Local Similarity   100.0%;  Pred. No. 4.4e-41;
Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          141 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLG 191
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db           48 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLG 98

RESULT 6
ABB67853
ID   ABB67853 standard; Protein; 311 AA.
XX
AC   ABB67853;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 30351.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11956.
```

```
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 30351; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC   sequences (ABL16176-ABL30511), expressed DNA
CC   (ABB5737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   311 AA;

Query Match        3.6%;    Score 13;   DB 22;   Length 311;
Best Local Similarity   100.0%;   Pred. No. 0.00062;
Matches  13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        324 GSSSSSSSSSTP 336
              ||||||||||||
Db        188 GSSSSSSSSSTP 200


RESULT 7
AAU32649
ID   AAU32649 standard; Protein; 106 AA.
XX
AC   AAU32649;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #3140.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 643; Page 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or

CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   106 AA;

Query Match        3.3%;    Score 12;   DB 22;   Length 106;
Best Local Similarity   100.0%;   Pred. No. 0.0021;
Matches  12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        323 KGSSSSSSSSS 334
              |||||||||||
Db         65 KGSSSSSSSSS 76


RESULT 8
ABB70839
ID   ABB70839 standard; Protein; 1262 AA.
XX
AC   ABB70839;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 39309.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916372.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL14942.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 39309; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB5737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1262 AA;
```

```
Query Match         3.3%;   Score 12;   DB 22;   Length 1262;
Best Local Similarity  100.0%;   Pred. No. 0.022;
Matches   12;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     324 GSSSSSSSSSST 335
           ||||||||||||
Db     351 GSSSSSSSSSST 362
```

RESULT 9
AAU32705
ID   AAU32705 standard; Protein; 85 AA.
XX
AC   AAU32705;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Novel human secreted protein #3196.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 652; Page 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   85 AA;

```
Query Match         3.1%;   Score 11;   DB 22;   Length 85;
Best Local Similarity  100.0%;   Pred. No. 0.016;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     324 GSSSSSSSSSS 334
           |||||||||||
Db      42 GSSSSSSSSSS 52
```

RESULT 10
ABP10952
ID   ABP10952 standard; Protein; 86 AA.
XX
AC   ABP10952;
XX
DT   24-JUN-2002 (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:21886.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN26704.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 21886; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B. The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   86 AA;

```
Query Match         3.1%;   Score 11;   DB 23;   Length 86;
```

```
Best Local Similarity  100.0%;  Pred. No. 0.016;
Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      324 GSSSSSSSSS 334
            ||||||||||
Db       27 GSSSSSSSSS 37


RESULT 11
AAU16320
ID   AAU16320 standard; Protein;  94 AA.
XX
AC   AAU16320;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1273.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic;
KW   cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmological;
KW   vulnerary; secreted protein; rheumatoid arthritis;
KW   hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW   cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW   corneal infection; wound healing; epithelial cell proliferation;
KW   skin ageing; food additive; preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01341.
XX
PR   31-JAN-2000; 2000US-0179065.        PR   01-SEP-2000; 2000US-0229343.
PR   04-FEB-2000; 2000US-0180628.        PR   01-SEP-2000; 2000US-0229344.
PR   24-FEB-2000; 2000US-0184464.        PR   01-SEP-2000; 2000US-0229345.
PR   02-MAR-2000; 2000US-0186350.        PR   05-SEP-2000; 2000US-0229509.
PR   16-MAR-2000; 2000US-0189874.        PR   05-SEP-2000; 2000US-0229513.
PR   17-MAR-2000; 2000US-0190076.        PR   06-SEP-2000; 2000US-0230437.
PR   18-APR-2000; 2000US-0198123.        PR   06-SEP-2000; 2000US-0230438.
PR   19-MAY-2000; 2000US-0205515.        PR   08-SEP-2000; 2000US-0231242.
PR   07-JUN-2000; 2000US-0209467.        PR   08-SEP-2000; 2000US-0231243.
PR   28-JUN-2000; 2000US-0214886.        PR   08-SEP-2000; 2000US-0231244.
PR   30-JUN-2000; 2000US-0215135.        PR   08-SEP-2000; 2000US-0231413.
PR   07-JUL-2000; 2000US-0216647.        PR   08-SEP-2000; 2000US-0231414.
PR   07-JUL-2000; 2000US-0216880.        PR   08-SEP-2000; 2000US-0232080.
PR   11-JUL-2000; 2000US-0217487.        PR   08-SEP-2000; 2000US-0232081.
PR   11-JUL-2000; 2000US-0217496.        PR   12-SEP-2000; 2000US-0231968.
PR   14-JUL-2000; 2000US-0218290.        PR   14-SEP-2000; 2000US-0232397.
PR   26-JUL-2000; 2000US-0220963.        PR   14-SEP-2000; 2000US-0232398.
PR   26-JUL-2000; 2000US-0220964.        PR   14-SEP-2000; 2000US-0232399.
PR   14-AUG-2000; 2000US-0224518.        PR   14-SEP-2000; 2000US-0232400.
PR   14-AUG-2000; 2000US-0224519.        PR   14-SEP-2000; 2000US-0232401.
PR   14-AUG-2000; 2000US-0225213.        PR   14-SEP-2000; 2000US-0233063.
PR   14-AUG-2000; 2000US-0225214.        PR   14-SEP-2000; 2000US-0233064.
PR   14-AUG-2000; 2000US-0225266.        PR   14-SEP-2000; 2000US-0233065.
PR   14-AUG-2000; 2000US-0225267.        PR   21-SEP-2000; 2000US-0234223.
PR   14-AUG-2000; 2000US-0225268.        PR   21-SEP-2000; 2000US-0234274.
PR   14-AUG-2000; 2000US-0225270.        PR   25-SEP-2000; 2000US-0234997.
PR   14-AUG-2000; 2000US-0225447.        PR   25-SEP-2000; 2000US-0234998.
PR   14-AUG-2000; 2000US-0225757.        PR   26-SEP-2000; 2000US-0235484.
PR   14-AUG-2000; 2000US-0225758.        PR   27-SEP-2000; 2000US-0235834.
PR   14-AUG-2000; 2000US-0225759.        PR   27-SEP-2000; 2000US-0235836.
PR   18-AUG-2000; 2000US-0226279.        PR   29-SEP-2000; 2000US-0236327.
PR   22-AUG-2000; 2000US-0226681.        PR   29-SEP-2000; 2000US-0236367.
PR   22-AUG-2000; 2000US-0226868.        PR   29-SEP-2000; 2000US-0236368.
PR   22-AUG-2000; 2000US-0227009.        PR   29-SEP-2000; 2000US-0236369.
PR   23-AUG-2000; 2000US-0227182.        PR   29-SEP-2000; 2000US-0236370.
PR   30-AUG-2000; 2000US-0228924.        PR   02-OCT-2000; 2000US-0236802.
PR   01-SEP-2000; 2000US-0229287.        PR   02-OCT-2000; 2000US-0237037.
                                         PR   02-OCT-2000; 2000US-0237038.
                                         PR   02-OCT-2000; 2000US-0237039.
                                         PR   02-OCT-2000; 2000US-0237040.
                                         PR   13-OCT-2000; 2000US-0239935.
                                         PR   13-OCT-2000; 2000US-0239937.
                                         PR   20-OCT-2000; 2000US-0240960.
                                         PR   20-OCT-2000; 2000US-0241221.
                                         PR   20-OCT-2000; 2000US-0241785.
                                         PR   20-OCT-2000; 2000US-0241786.
                                         PR   20-OCT-2000; 2000US-0241787.
                                         PR   20-OCT-2000; 2000US-0241808.
                                         PR   20-OCT-2000; 2000US-0241809.
                                         PR   01-NOV-2000; 2000US-0244617.
                                         PR   08-NOV-2000; 2000US-0246474.
                                         PR   08-NOV-2000; 2000US-0246475.
                                         PR   08-NOV-2000; 2000US-0246476.
                                         PR   08-NOV-2000; 2000US-0246477.
                                         PR   08-NOV-2000; 2000US-0246478.
                                         PR   08-NOV-2000; 2000US-0246523.
                                         PR   08-NOV-2000; 2000US-0246524.
                                         PR   08-NOV-2000; 2000US-0246525.
                                         PR   08-NOV-2000; 2000US-0246526.
                                         PR   08-NOV-2000; 2000US-0246527.
                                         PR   08-NOV-2000; 2000US-0246528.
                                         PR   08-NOV-2000; 2000US-0246532.
                                         PR   08-NOV-2000; 2000US-0246609.
                                         PR   08-NOV-2000; 2000US-0246610.
                                         PR   08-NOV-2000; 2000US-0246611.
                                         PR   08-NOV-2000; 2000US-0246613.
                                         PR   17-NOV-2000; 2000US-0249207.
                                         PR   17-NOV-2000; 2000US-0249208.
                                         PR   17-NOV-2000; 2000US-0249209.
                                         PR   17-NOV-2000; 2000US-0249210.
                                         PR   17-NOV-2000; 2000US-0249211.
                                         PR   17-NOV-2000; 2000US-0249212.
```

```
RESULT 12
ABU55389
ID   ABU55389 standard; Protein; 94 AA.
XX
AC   ABU55389;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #476.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US200213753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-0764864.
XX
PR   31-JAN-2000; 2000US-1790065P.
PR   04-FEB-2000; 2000US-1806628P.
PR   28-JUN-2000; 2000US-2148868P.
PR   07-JUL-2000; 2000US-2166447P.
PR   07-JUL-2000; 2000US-2168808P.
PR   11-JUL-2000; 2000US-2174887P.
PR   11-JUL-2000; 2000US-2174962P.
PR   14-JUL-2000; 2000US-2182902P.
PR   26-JUL-2000; 2000US-2209663P.
PR   26-JUL-2000; 2000US-2209664P.
PR   14-AUG-2000; 2000US-2245188P.
PR   14-AUG-2000; 2000US-2245192P.
PR   14-AUG-2000; 2000US-2252677P.
PR   14-AUG-2000; 2000US-2252688P.
PR   14-AUG-2000; 2000US-2252702P.
PR   14-AUG-2000; 2000US-2254477P.
PR   14-AUG-2000; 2000US-2255778P.
PR   14-AUG-2000; 2000US-2257588P.
PR   22-AUG-2000; 2000US-2268688P.
PR   30-AUG-2000; 2000US-2289244P.
PR   01-SEP-2000; 2000US-2292877P.
PR   01-SEP-2000; 2000US-2293438P.
PR   01-SEP-2000; 2000US-2293448P.
PR   01-SEP-2000; 2000US-2293458P.
PR   05-SEP-2000; 2000US-2295098P.
PR   05-SEP-2000; 2000US-2295138P.
PR   08-SEP-2000; 2000US-2311413P.
PR   21-SEP-2000; 2000US-2342238P.
PR   21-SEP-2000; 2000US-2342748P.
PR   25-SEP-2000; 2000US-2349978P.
PR   27-SEP-2000; 2000US-2358348P.
PR   29-SEP-2000; 2000US-2363278P.
PR   29-SEP-2000; 2000US-2363678P.
PR   29-SEP-2000; 2000US-2363688P.
PR   29-SEP-2000; 2000US-2363698P.
PR   29-SEP-2000; 2000US-2363708P.
PR   02-OCT-2000; 2000US-2368028P.
PR   02-OCT-2000; 2000US-2370378P.
PR   02-OCT-2000; 2000US-2370388P.
PR   02-OCT-2000; 2000US-2370398P.
PR   02-OCT-2000; 2000US-2370408P.
PR   13-OCT-2000; 2000US-2399358P.
PR   20-OCT-2000; 2000US-2409608P.
PR   20-OCT-2000; 2000US-2417858P.
PR   20-OCT-2000; 2000US-2418098P.
PR   01-NOV-2000; 2000US-2446178P.
PR   17-NOV-2000; 2000US-2492138.
PR   17-NOV-2000; 2000US-2492148.
PR   17-NOV-2000; 2000US-2492158.
PR   17-NOV-2000; 2000US-2492168.
PR   17-NOV-2000; 2000US-2492178.
PR   17-NOV-2000; 2000US-2492188.
PR   17-NOV-2000; 2000US-2492448.
PR   17-NOV-2000; 2000US-2492458.
PR   17-NOV-2000; 2000US-2492648.
PR   17-NOV-2000; 2000US-2492658.
PR   17-NOV-2000; 2000US-2492978.
PR   17-NOV-2000; 2000US-2492998.
PR   17-NOV-2000; 2000US-2493008.
PR   01-DEC-2000; 2000US-2501608.
PR   01-DEC-2000; 2000US-2503918.
PR   05-DEC-2000; 2000US-2510308.
PR   05-DEC-2000; 2000US-2519888.
PR   05-DEC-2000; 2000US-2567198.
PR   06-DEC-2000; 2000US-2514798.
PR   08-DEC-2000; 2000US-2518568.
PR   08-DEC-2000; 2000US-2516688.
PR   08-DEC-2000; 2000US-2518698.
PR   08-DEC-2000; 2000US-2519898.
PR   08-DEC-2000; 2000US-2519908.
PR   11-DEC-2000; 2000US-2540978.
PR   05-JAN-2001; 2001US-2596788.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26307.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 11; SEQ ID No 1273; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC   are also used in diagnosing a pathological condition or susceptibility
CC   to a pathological condition. Antibodies to the proteins can also
CC   be used in alleviating symptoms associated with the disorders and in
CC   diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC   immunosorbant assays (ELISA). Disorders which are diagnosed or treated
CC   include autoimmune diseases e.g. rheumatoid arthritis,
CC   hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC   cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC   e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC   Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC   and ocular disorders e.g. corneal infection, and many other
CC   disorders listed in the specification. The polypeptides can also
CC   be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention.

Query Match          3.1%;  Score 11;  DB 22;  Length 94;
Best Local Similarity  100.0%;  Pred. No. 0.018;
Matches   11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSSS 334
          |||||||||||
Db     17 GSSSSSSSSSS 27
```

PR   17-NOV-2000; 2000US-249299P.
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73648.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for
PT   treating, inhibiting or preventing e.g. neural, immune system,
PT   muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT   cardiovascular or renal disorders -
XX
PS   Claim 11; SEQ ID NO 1273; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC   arthritis and multiple sclerosis), muscular disorders, respiratory
CC   diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC   reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC   cardiovascular disorders (e.g. congenital heart defects, Ebstein's
CC   anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC   kidney failure and end-stage renal disease), hyperproliferative disorders
CC   (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.
CC   septic shock, bursitis and appendicitis), allergic reactions and
CC   conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC   atherosclerosis and myocardial infarction) and cancerous diseases.
CC   Sequences ABU54914-ABU55669 and ABU55748 represent human novel
CC   polypeptides of the invention.
XX
SQ   Sequence   94 AA;

Query Match   3.1%; Score 11; DB 24; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
               |||||||||||
Db   17 GSSSSSSSSS 27

RESULT 13
ABG58455
ID   ABG58455 standard; Peptide; 104 AA.
XX
AC   ABG58455;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human liver peptide, SEQ ID NO 37103.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00664.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.

PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analysing gene expression in human adult liver -
XX
PS   Claim 27; SEQ ID No 37103; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult
CC   liver. (I) may be used for predicting, measuring and displaying gene
CC   expression in samples derived from human adult liver. The genes
CC   identified may be involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC   is associated with coronary heart disease. ABG47348-ABG59930 represent
CC   human liver single exon encoded peptides of the invention.
CC   Note: The sequence information for this patent does not appear in the
CC   printed specification but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   104 AA;

Query Match   3.1%; Score 11; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSST 335
               |||||||||||
Db   83 SSSSSSSSSST 93

RESULT 14
ABB43053
ID   ABB43053 standard; Peptide; 104 AA.
XX
AC   ABB43053;
XX
DT   04-FEB-2002 (first entry)
XX
DE   Peptide #10559 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 27; SEQ ID NO 35688; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a peptide encoded by a single exon
CC   nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   104 AA;

Query Match   3.1%; Score 11; DB 22; Length 104;
Best Local Similarity   100.0%; Pred. No. 0.02;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSST 335
          ||||||||||
Db   83 SSSSSSSSSST 93

RESULT 15
ABB26224
ID   ABB26224 standard; Protein; 104 AA.
XX
AC   ABB26224;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Protein #8223 encoded by probe for measuring heart cell gene expression.
XX
KW   Human; gene expression; heart; microarray; vascular system;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 15; SEQ ID No 27994; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart (see
CC   ABA21535-ABA41305). The present sequence is a protein encoded by one such

CC   probe. The probes may be used for predicting, measuring and displaying
CC   gene expression in samples derived from the human heart via microarrays.
CC   By measuring gene expression, the probes are useful for predicting,
CC   diagnosing, grading, staging, monitoring and prognosing diseases of the
CC   human heart and vascular system e.g. cardiovascular disease,
CC   hypertension, cardiac arrhythmias and congenital heart disease.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   104 AA;

Query Match   3.1%; Score 11; DB 22; Length 104;
Best Local Similarity   100.0%; Pred. No. 0.02;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSST 335
          ||||||||||
Db   83 SSSSSSSSSST 93

Search completed: February 19, 2004, 19:37:18
Job time : 44 secs

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February 19, 2004, 19:36:26 ; Search time 21 Seconds
                                 (without alignments)
                                 725.329 Million cell updates/sec

Title:         US-10-041-615-34
Perfect score: 360
Sequence:      1 MIKLJGPYFTPPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched:      328717 seqs, 42310858 residues

Word size :    10

Total number of hits satisfying chosen parameters:      97

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   Issued Patents AA:*
   1: /cgm2_6/ptodata/1/iaa/5A_COMB.pep:*
   2: /cgm2_6/ptodata/1/iaa/5B_COMB.pep:*
   3: /cgm2_6/ptodata/1/iaa/6A_COMB.pep:*
   4: /cgm2_6/ptodata/1/iaa/6B_COMB.pep:*
   5: /cgm2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
   6: /cgm2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	11	3.1	43	4	US-07-757-022B-12	Sequence 12, Appl
2	11	3.1	192	4	US-07-757-022B-90	Sequence 90, Appl
3	11	3.1	204	4	US-07-757-022B-92	Sequence 92, Appl
4	11	3.1	208	4	US-07-757-022B-132	Sequence 132, App
5	11	3.1	209	4	US-07-757-022B-94	Sequence 94, Appl
6	11	3.1	217	4	US-07-757-022B-76	Sequence 76, Appl
7	11	3.1	220	4	US-07-757-022B-96	Sequence 96, Appl
8	11	3.1	231	4	US-07-757-022B-30	Sequence 30, Appl
9	11	3.1	296	4	US-07-757-022B-70	Sequence 70, Appl
10	11	3.1	310	3	US-08-651-136C-22	Sequence 22, Appl
11	11	3.1	310	4	US-09-229-911A-22	Sequence 22, Appl
12	11	3.1	372	4	US-07-757-022B-64	Sequence 64, Appl
13	11	3.1	422	4	US-07-757-022B-68	Sequence 68, Appl
14	11	3.1	423	4	US-07-757-022B-66	Sequence 66, Appl
15	11	3.1	463	4	US-07-757-022B-54	Sequence 54, Appl
16	11	3.1	551	4	US-09-252-991A-27787	Sequence 27787, A
17	11	3.1	568	1	US-08-320-559-30	Sequence 30, Appl
18	11	3.1	568	5	US-08-545-860D-30	Sequence 30, Appl
19	11	3.1	568	5	PCT-US94-04496-30	Sequence 30, Appl
20	11	3.1	609	4	US-09-198-452A-579	Sequence 579, App
21	11	3.1	693	4	US-09-107-532A-5812	Sequence 5812, AP
22	11	3.1	703	3	US-08-910-925-4	Sequence 4, Appli
23	11	3.1	717	3	US-08-910-925-1	Sequence 1, Appli
24	11	3.1	743	3	US-08-910-925-3	Sequence 3, Appli
25	11	3.1	1049	4	US-07-757-022B-58	Sequence 58, Appl
26	11	3.1	1140	4	US-07-757-022B-104	Sequence 104, App
27	11	3.1	1313	4	US-07-757-022B-142	Sequence 142, App
28	11	3.1	1314	4	US-07-757-022B-50	Sequence 50, Appl
29	11	3.1	1354	4	US-07-757-022B-48	Sequence 48, Appl
30	11	3.1	1363	4	US-07-757-022B-52	Sequence 52, Appl
31	11	3.1	1404	4	US-07-757-022B-2	Sequence 2, Appl
32	11	3.1	1404	4	US-07-757-022B-62	Sequence 62, Appl
33	10	2.8	130	2	US-08-630-822A-90	Sequence 90, Appl
34	10	2.8	130	2	US-09-005-069-90	Sequence 90, Appl
35	10	2.8	130	4	US-09-171-156A-39	Sequence 39, Appl
36	10	2.8	130	4	US-09-004-730A-39	Sequence 39, Appl
37	10	2.8	130	4	US-08-981-799A-39	Sequence 39, Appl
38	10	2.8	148	2	US-09-461-325-453	Sequence 453, App
39	10	2.8	218	3	US-09-068-655-7	Sequence 7, Appli
40	10	2.8	223	4	US-09-009-816-4	Sequence 4, Appli
41	10	2.8	229	4	US-09-604-978-9	Sequence 9, Appli
42	10	2.8	229	4	US-09-604-728-9	Sequence 9, Appli
43	10	2.8	349	4	US-09-162-524-3	Sequence 3, Appli
44	10	2.8	367	4	US-08-213-419B-15	Sequence 15, Appli
45	10	2.8	367	4	US-09-009-816-2	Sequence 2, Appli

```
                        ALIGNMENTS

RESULT 1
US-07-757-022B-12
; Sequence 12, Application US/07757022B
; Patent No. 643142
; GENERAL INFORMATION:
;  APPLICANT: Gesner, Thomas G.
;  APPLICANT: Clark, Stephen C.
;  APPLICANT: Turner, Katherine
;  APPLICANT: Hewick, Rodney M.
;  TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;  NUMBER OF SEQUENCES: 143
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genetics Institute, Inc.
;   STREET: 87 CambridgePark Drive
;   CITY: Cambridge
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02140
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/757,022B
;   FILING DATE: 19910910
;   CLASSIFICATION: 530
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/643,502
;   FILING DATE: 18-JAN-1991
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/546,114
;   FILING DATE: 29-JUN-1990
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/457,196
;   FILING DATE: 29-DEC-1989
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/390,901
;   FILING DATE: 08-AUG-1989
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Cserr, Luann
;   REGISTRATION NUMBER: 31,822
;   REFERENCE/DOCKET NUMBER: GI 5190
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)876-1170
;   TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 12:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 43 amino acids
;   TYPE: AMINO ACID
```

```
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-12

Query Match     3.1%;  Score 11;  DB 4;  Length 43;
Best Local Similarity  100.0%;  Pred. No. 0.0038;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          |||||||||||
Db     12 SSSSSSSSST 22

RESULT 2
US-07-757-022B-90
; Sequence 90, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
;   APPLICANT: Gesner, Thomas G.
;   APPLICANT: Clark, Stephen C.
;   APPLICANT: Turner, Katherine
;   APPLICANT: Hewick, Rodney M.
;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;   NUMBER OF SEQUENCES: 143
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genetics Institute, Inc.
;     STREET: 87 CambridgePark Drive
;     CITY: Cambridge
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 02140
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/757,022B
;     FILING DATE: 19910910
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/643,502
;     FILING DATE: 18-JAN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/546,114
;     FILING DATE: 29-JUN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/457,196
;     FILING DATE: 29-DEC-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/390,901
;     FILING DATE: 08-AUG-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Cserr, Luann
;     REGISTRATION NUMBER: 31,822
;     REFERENCE/DOCKET NUMBER: GI 5190
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617)876-1170
;     TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 90:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 192 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-07-757-022B-90

Query Match     3.1%;  Score 11;  DB 4;  Length 192;
Best Local Similarity  100.0%;  Pred. No. 0.015;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          |||||||||||

Db    168 SSSSSSSSST 178

RESULT 3
US-07-757-022B-92
; Sequence 92, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
;   APPLICANT: Gesner, Thomas G.
;   APPLICANT: Clark, Stephen C.
;   APPLICANT: Turner, Katherine
;   APPLICANT: Hewick, Rodney M.
;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;   NUMBER OF SEQUENCES: 143
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genetics Institute, Inc.
;     STREET: 87 CambridgePark Drive
;     CITY: Cambridge
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 02140
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/757,022B
;     FILING DATE: 19910910
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/643,502
;     FILING DATE: 18-JAN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/546,114
;     FILING DATE: 29-JUN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/457,196
;     FILING DATE: 29-DEC-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/390,901
;     FILING DATE: 08-AUG-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Cserr, Luann
;     REGISTRATION NUMBER: 31,822
;     REFERENCE/DOCKET NUMBER: GI 5190
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617)876-1170
;     TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 92:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 204 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-07-757-022B-92

Query Match     3.1%;  Score 11;  DB 4;  Length 204;
Best Local Similarity  100.0%;  Pred. No. 0.016;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          |||||||||||
Db    168 SSSSSSSSST 178

RESULT 4
US-07-757-022B-132
; Sequence 132, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
;   APPLICANT: Gesner, Thomas G.
;   APPLICANT: Clark, Stephen C.
```

```
 ;   APPLICANT: Turner, Katherine
 ;   APPLICANT: Hewick, Rodney M.
 ;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ;   NUMBER OF SEQUENCES: 143
 ;   CORRESPONDENCE ADDRESS:
 ;     ADDRESSEE: Genetics Institute, Inc.
 ;     STREET: 87 CambridgePark Drive
 ;     CITY: Cambridge
 ;     STATE: Massachusetts
 ;     COUNTRY: U.S.A.
 ;     ZIP: 02140
 ;   COMPUTER READABLE FORM:
 ;     MEDIUM TYPE: Floppy disk
 ;     COMPUTER: IBM PC compatible
 ;     OPERATING SYSTEM: PC-DOS/MS-DOS
 ;     SOFTWARE: PatentIn Release #1.0, Version #1.25
 ;   CURRENT APPLICATION DATA:
 ;     APPLICATION NUMBER: US/07/757,022B
 ;     FILING DATE: 19910910
 ;     CLASSIFICATION: 530
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/643,502
 ;     FILING DATE: 18-JAN-1991
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/546,114
 ;     FILING DATE: 29-JUN-1990
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/457,196
 ;     FILING DATE: 29-DEC-1989
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/390,901
 ;     FILING DATE: 08-AUG-1989
 ;   ATTORNEY/AGENT INFORMATION:
 ;     NAME: Cserr, Luann
 ;     REGISTRATION NUMBER: 31,822
 ;     REFERENCE/DOCKET NUMBER: GI 5190
 ;   TELECOMMUNICATION INFORMATION:
 ;     TELEPHONE: (617)876-1170
 ;     TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 132:
 ;   SEQUENCE CHARACTERISTICS:
 ;     LENGTH: 208 amino acids
 ;     TYPE: AMINO ACID
 ;     TOPOLOGY: linear
 ;   MOLECULE TYPE: protein
US-07-757-022B-132

Query Match          3.1%;   Score 11;   DB 4;   Length 208;
Best Local Similarity   100.0%;   Pred. No. 0.016;   Indels   0;   Gaps   0;
Matches   11;   Conservative   0;   Mismatches   0;

Qy     325 SSSSSSSSST 335
           |||||||||||
Db     168 SSSSSSSSST 178

RESULT 5
US-07-757-022B-94
; Sequence 94, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
 ;   APPLICANT: Gesner, Thomas G.
 ;   APPLICANT: Clark, Stephen C.
 ;   APPLICANT: Turner, Katherine
 ;   APPLICANT: Hewick, Rodney M.
 ;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ;   NUMBER OF SEQUENCES: 143
 ;   CORRESPONDENCE ADDRESS:
 ;     ADDRESSEE: Genetics Institute, Inc.
 ;     STREET: 87 CambridgePark Drive
 ;     CITY: Cambridge
 ;     STATE: Massachusetts
 ;     COUNTRY: U.S.A.
```

```
 ;     ZIP: 02140
 ;   COMPUTER READABLE FORM:
 ;     MEDIUM TYPE: Floppy disk
 ;     COMPUTER: IBM PC compatible
 ;     OPERATING SYSTEM: PC-DOS/MS-DOS
 ;     SOFTWARE: PatentIn Release #1.0, Version #1.25
 ;   CURRENT APPLICATION DATA:
 ;     APPLICATION NUMBER: US/07/757,022B
 ;     FILING DATE: 19910910
 ;     CLASSIFICATION: 530
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/643,502
 ;     FILING DATE: 18-JAN-1991
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/546,114
 ;     FILING DATE: 29-JUN-1990
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/457,196
 ;     FILING DATE: 29-DEC-1989
 ;   PRIOR APPLICATION DATA:
 ;     APPLICATION NUMBER: US 07/390,901
 ;     FILING DATE: 08-AUG-1989
 ;   ATTORNEY/AGENT INFORMATION:
 ;     NAME: Cserr, Luann
 ;     REGISTRATION NUMBER: 31,822
 ;     REFERENCE/DOCKET NUMBER: GI 5190
 ;   TELECOMMUNICATION INFORMATION:
 ;     TELEPHONE: (617)876-1170
 ;     TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 94:
 ;   SEQUENCE CHARACTERISTICS:
 ;     LENGTH: 209 amino acids
 ;     TYPE: AMINO ACID
 ;     TOPOLOGY: linear
 ;   MOLECULE TYPE: protein
US-07-757-022B-94

Query Match          3.1%;   Score 11;   DB 4;   Length 209;
Best Local Similarity   100.0%;   Pred. No. 0.016;   Indels   0;   Gaps   0;
Matches   11;   Conservative   0;   Mismatches   0;

Qy     325 SSSSSSSSST 335
           |||||||||||
Db     168 SSSSSSSSST 178

RESULT 6
US-07-757-022B-76
; Sequence 76, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
 ;   APPLICANT: Gesner, Thomas G.
 ;   APPLICANT: Clark, Stephen C.
 ;   APPLICANT: Turner, Katherine
 ;   APPLICANT: Hewick, Rodney M.
 ;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ;   NUMBER OF SEQUENCES: 143
 ;   CORRESPONDENCE ADDRESS:
 ;     ADDRESSEE: Genetics Institute, Inc.
 ;     STREET: 87 CambridgePark Drive
 ;     CITY: Cambridge
 ;     STATE: Massachusetts
 ;     COUNTRY: U.S.A.
 ;     ZIP: 02140
 ;   COMPUTER READABLE FORM:
 ;     MEDIUM TYPE: Floppy disk
 ;     COMPUTER: IBM PC compatible
 ;     OPERATING SYSTEM: PC-DOS/MS-DOS
 ;     SOFTWARE: PatentIn Release #1.0, Version #1.25
 ;   CURRENT APPLICATION DATA:
 ;     APPLICATION NUMBER: US/07/757,022B
 ;     FILING DATE: 19910910
 ;     CLASSIFICATION: 530
```

```
       PRIOR APPLICATION DATA:
           APPLICATION NUMBER: US 07/643,502
           FILING DATE: 18-JAN-1991
       PRIOR APPLICATION DATA:
           APPLICATION NUMBER: US 07/546,114
           FILING DATE: 29-JUN-1990
       PRIOR APPLICATION DATA:
           APPLICATION NUMBER: US 07/457,196
           FILING DATE: 29-DEC-1989
       PRIOR APPLICATION DATA:
           APPLICATION NUMBER: US 07/390,901
           FILING DATE: 08-AUG-1989
       ATTORNEY/AGENT INFORMATION:
           NAME: Cserr, Luann
           REGISTRATION NUMBER: 31,822
           REFERENCE/DOCKET NUMBER: GI 5190
       TELECOMMUNICATION INFORMATION:
           TELEPHONE: (617)876-1170
           TELEFAX: (617)876-5851
   INFORMATION FOR SEQ ID NO: 76:
       SEQUENCE CHARACTERISTICS:
           LENGTH: 217 amino acids
           TYPE: AMINO ACID
           TOPOLOGY: linear
       MOLECULE TYPE: protein
US-07-757-022B-76

Query Match      3.1%;  Score 11;  DB 4;  Length 217;
Best Local Similarity   100.0%;  Pred. No. 0.017;
Matches   11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    325 SSSSSSSSSST 335
          ||||||||||
Db    128 SSSSSSSSSST 138

RESULT 7
US-07-757-022B-96
; Sequence 96, Application US/07757022B
; Patent No. 6433142
GENERAL INFORMATION:
    APPLICANT: Gesner, Thomas G.
    APPLICANT: Clark, Stephen C.
    APPLICANT: Turner, Katherine
    APPLICANT: Hewick, Rodney M.
    TITLE OF INVENTION: Megakaryocyte Stimulating Factors
    NUMBER OF SEQUENCES: 143
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Genetics Institute, Inc.
        STREET: 87 CambridgePark Drive
        CITY: Cambridge
        STATE: Massachusetts
        COUNTRY: U.S.A.
        ZIP: 02140
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/07/757,022B
        FILING DATE: 19910910
        CLASSIFICATION: 530
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/643,502
        FILING DATE: 18-JAN-1991
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/546,114
        FILING DATE: 29-JUN-1990
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/457,196
        FILING DATE: 29-DEC-1989
    PRIOR APPLICATION DATA:
```

```
           APPLICATION NUMBER: US 07/390,901
           FILING DATE: 08-AUG-1989
       ATTORNEY/AGENT INFORMATION:
           NAME: Cserr, Luann
           REGISTRATION NUMBER: 31,822
           REFERENCE/DOCKET NUMBER: GI 5190
       TELECOMMUNICATION INFORMATION:
           TELEPHONE: (617)876-1170
           TELEFAX: (617)876-5851
   INFORMATION FOR SEQ ID NO: 96:
       SEQUENCE CHARACTERISTICS:
           LENGTH: 220 amino acids
           TYPE: AMINO ACID
           TOPOLOGY: linear
       MOLECULE TYPE: protein
US-07-757-022B-96

Query Match      3.1%;  Score 11;  DB 4;  Length 220;
Best Local Similarity   100.0%;  Pred. No. 0.017;
Matches   11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    325 SSSSSSSSSST 335
          ||||||||||
Db    168 SSSSSSSSSST 178

RESULT 8
US-07-757-022B-30
; Sequence 30, Application US/07757022B
; Patent No. 6433142
GENERAL INFORMATION:
    APPLICANT: Gesner, Thomas G.
    APPLICANT: Clark, Stephen C.
    APPLICANT: Turner, Katherine
    APPLICANT: Hewick, Rodney M.
    TITLE OF INVENTION: Megakaryocyte Stimulating Factors
    NUMBER OF SEQUENCES: 143
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Genetics Institute, Inc.
        STREET: 87 CambridgePark Drive
        CITY: Cambridge
        STATE: Massachusetts
        COUNTRY: U.S.A.
        ZIP: 02140
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/07/757,022B
        FILING DATE: 19910910
        CLASSIFICATION: 530
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/643,502
        FILING DATE: 18-JAN-1991
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/546,114
        FILING DATE: 29-JUN-1990
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/457,196
        FILING DATE: 29-DEC-1989
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/390,901
        FILING DATE: 08-AUG-1989
    ATTORNEY/AGENT INFORMATION:
        NAME: Cserr, Luann
        REGISTRATION NUMBER: 31,822
        REFERENCE/DOCKET NUMBER: GI 5190
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (617)876-1170
        TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 30:
```

```
       SEQUENCE CHARACTERISTICS:
         LENGTH: 231 amino acids
         TYPE: AMINO ACID
         TOPOLOGY: linear
         MOLECULE TYPE: protein
US-07-757-022B-30

Query Match          3.1%;  Score 11;  DB 4;  Length 231;
Best Local Similarity   100.0%;  Pred. No. 0.018;
Matches  11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      325 SSSSSSSSST 335
            |||||||||||
Db      168 SSSSSSSSST 178


RESULT 9
US-07-757-022B-70
; Sequence 70, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
;   APPLICANT: Gesner, Thomas G.
;   APPLICANT: Clark, Stephen C.
;   APPLICANT: Turner, Katherine
;   APPLICANT: Hewick, Rodney M.
;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;   NUMBER OF SEQUENCES: 143
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genetics Institute, Inc.
;     STREET: 87 CambridgePark Drive
;     CITY: Cambridge
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 02140
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/07/757,022B
;     FILING DATE: 19910910
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/643,502
;     FILING DATE: 18-JAN-1991
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/546,114
;     FILING DATE: 29-JUN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/457,196
;     FILING DATE: 29-DEC-1989
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/390,901
;     FILING DATE: 08-AUG-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Cserr, Luann
;     REGISTRATION NUMBER: 31,822
;     REFERENCE/DOCKET NUMBER: GI 5190
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (617)876-1170
;     TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 70:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 296 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-07-757-022B-70

Query Match          3.1%;  Score 11;  DB 4;  Length 296;
Best Local Similarity   100.0%;  Pred. No. 0.022;
Matches  11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy      325 SSSSSSSSST 335
            |||||||||||
Db      168 SSSSSSSSST 178


RESULT 10
US-08-651-136C-22
; Sequence 22, Application US/08651136C
; Patent No. 6001639
; GENERAL INFORMATION:
;   APPLICANT: Schulein, Martin
;   APPLICANT: Andersen, Lene N.
;   APPLICANT: Lassen, Soren F.
;   APPLICANT: Kauppinen, Markus S.
;   APPLICANT: Lange, Lene
;   APPLICANT: Nielsen, Ruby I.
;   APPLICANT: Ihara, Michiko
;   APPLICANT: Takagi, Shinobu
;   TITLE OF INVENTION: No. 600163 9el Endoglucanases
;   NUMBER OF SEQUENCES: 109
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: No. 600163 9o No. 600163 9disk of No. 600163 9th America, Inc.
;     STREET: 405 Lexington Avenue, 64th Floor
;     CITY: New York
;     STATE: New York
;     COUNTRY: United States of America
;     ZIP: 10174-6401
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/651,136C
;     FILING DATE: 21-MAY-1996
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Lambiris, Elias J.
;     REGISTRATION NUMBER: 33,728
;     REFERENCE/DOCKET NUMBER: 4366.200-US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-867-0123
;     TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO:     22:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 310 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-651-136C-22

Query Match          3.1%;  Score 11;  DB 3;  Length 310;
Best Local Similarity   100.0%;  Pred.No. 0.023;
Matches  11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      325 SSSSSSSSST 335
            |||||||||||
Db      248 SSSSSSSSST 258


RESULT 11
US-09-229-911A-22
; Sequence 22, Application US/09229911A
; Patent No. 6387690
; GENERAL INFORMATION:
;   APPLICANT: Schulein, Martin
;     Andersen, Lene N.
;     Lassen, Soren F.
;     Kauppinen, Markus S.
;     Lange, Lene
;     Nielsen, Ruby I.
;     Ihara, Michiko
```

Takagi, Shinobu
TITLE OF INVENTION: No. 6387690el Endoglucanases
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6387690o No. 6387690disk of No. 6387690th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/229,911A
FILING DATE: 13-Jan-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651,136
FILING DATE: 21-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4366.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-229-911A-22

Query Match   3.1%; Score 11; DB 4; Length 310;
Best Local Similarity   100.0%; Pred. No. 0.023;
Matches 11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSST 335
         |||||||||
Db   248 SSSSSSSSST 258

RESULT 12
US-07-757-022B-64
; Sequence 64, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-64

Query Match   3.1%; Score 11; DB 4; Length 372;
Best Local Similarity   100.0%; Pred. No. 0.027;
Matches 11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSST 335
         |||||||||
Db   77 SSSSSSSSST 87

RESULT 13
US-07-757-022B-68
; Sequence 68, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989

```
    PRIOR APPLICATION DATA:
     APPLICATION NUMBER: US 07/390,901
     FILING DATE: 08-AUG-1989
    ATTORNEY/AGENT INFORMATION:
     NAME: Cserr, Luann
     REGISTRATION NUMBER: 31,822
     REFERENCE/DOCKET NUMBER: GI 5190
    TELECOMMUNICATION INFORMATION:
     TELEPHONE: (617)876-1170
     TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 68:
    SEQUENCE CHARACTERISTICS:
     LENGTH: 422 amino acids
     TYPE: AMINO ACID
     TOPOLOGY: linear
     MOLECULE TYPE: protein
US-07-757-022B-68

Query Match         3.1%;  Score 11;  DB 4;  Length 422;
Best Local Similarity 100.0%;  Pred. No. 0.031;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  325 SSSSSSSSST 335
        |||||||||||
Db  127 SSSSSSSSSST 137

RESULT 14
US-07-757-022B-66
  Sequence 66, Application US/07757022B
  Patent No. 6433142
  GENERAL INFORMATION:
     APPLICANT: Gesner, Thomas G.
     APPLICANT: Clark, Stephen C.
     APPLICANT: Turner, Katherine
     APPLICANT: Hewick, Rodney M.
     TITLE OF INVENTION: Megakaryocyte Stimulating Factors
     NUMBER OF SEQUENCES: 143
     CORRESPONDENCE ADDRESS:
      ADDRESSEE: Genetics Institute, Inc.
      STREET: 87 CambridgePark Drive
      CITY: Cambridge
      STATE: Massachusetts
      COUNTRY: U.S.A.
      ZIP: 02140
     COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/07/757,022B
      FILING DATE: 19910910
      CLASSIFICATION: 530
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/643,502
      FILING DATE: 18-JAN-1991
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/546,114
      FILING DATE: 29-JUN-1990
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/457,196
      FILING DATE: 29-DEC-1989
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/390,901
      FILING DATE: 08-AUG-1989
     ATTORNEY/AGENT INFORMATION:
      NAME: Cserr, Luann
      REGISTRATION NUMBER: 31,822
      REFERENCE/DOCKET NUMBER: GI 5190
     TELECOMMUNICATION INFORMATION:
      TELEPHONE: (617)876-1170
      TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 66:
    SEQUENCE CHARACTERISTICS:
     LENGTH: 423 amino acids
     TYPE: AMINO ACID
     TOPOLOGY: linear
     MOLECULE TYPE: protein
US-07-757-022B-66

Query Match         3.1%;  Score 11;  DB 4;  Length 423;
Best Local Similarity 100.0%;  Pred. No. 0.031;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  325 SSSSSSSSST 335
        |||||||||||
Db  128 SSSSSSSSSST 138

RESULT 15
US-07-757-022B-54
  Sequence 54, Application US/07757022B
  Patent No. 6433142
  GENERAL INFORMATION:
     APPLICANT: Gesner, Thomas G.
     APPLICANT: Clark, Stephen C.
     APPLICANT: Turner, Katherine
     APPLICANT: Hewick, Rodney M.
     TITLE OF INVENTION: Megakaryocyte Stimulating Factors
     NUMBER OF SEQUENCES: 143
     CORRESPONDENCE ADDRESS:
      ADDRESSEE: Genetics Institute, Inc.
      STREET: 87 CambridgePark Drive
      CITY: Cambridge
      STATE: Massachusetts
      COUNTRY: U.S.A.
      ZIP: 02140
     COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/07/757,022B
      FILING DATE: 19910910
      CLASSIFICATION: 530
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/643,502
      FILING DATE: 18-JAN-1991
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/546,114
      FILING DATE: 29-JUN-1990
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/457,196
      FILING DATE: 29-DEC-1989
     PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/390,901
      FILING DATE: 08-AUG-1989
     ATTORNEY/AGENT INFORMATION:
      NAME: Cserr, Luann
      REGISTRATION NUMBER: 31,822
      REFERENCE/DOCKET NUMBER: GI 5190
     TELECOMMUNICATION INFORMATION:
      TELEPHONE: (617)876-1170
      TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 54:
    SEQUENCE CHARACTERISTICS:
     LENGTH: 463 amino acids
     TYPE: AMINO ACID
     TOPOLOGY: linear
     MOLECULE TYPE: protein
US-07-757-022B-54

Query Match         3.1%;  Score 11;  DB 4;  Length 463;
Best Local Similarity 100.0%;  Pred. No. 0.033;
```

Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   325 SSSSSSSSSST 335
         ||||||||||
Db   168 SSSSSSSSSST 178
```

Search completed: February 19, 2004, 19:39:48
Job time : 22 secs

```
                 GenCore version 5.1.6

         Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 19, 2004, 19:37:21 ; Search time 40 Seconds
                                         (without alignments)
                                         1884.441 Million cell updates/sec

Title:      US-10-041-615-34
Perfect score: 360
Sequence:   1 MIKLGPYFTPPTKIKTKIKD..........SETPSITQARGSMFLAEHVV 360

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched:   801455 seqs, 209382283 residues

Word size : 10

Total number of hits satisfying chosen parameters:      763

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  Published Applications AA:*
  1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
  2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
  3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
  4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
  5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
  6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
  7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
  8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
  9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
 15: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                           SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	360	100.0	360	12	US-10-041-615-34	Sequence 34, Appl
2	110	30.6	229	12	US-10-041-615-109	Sequence 109, App
3	110	30.6	323	12	US-10-017-161-1620	Sequence 1620, Ap
4	110	30.6	323	12	US-10-292-798-1294	Sequence 1294, Ap
5	70	19.4	177	12	US-09-801-944B-198	Sequence 198, Appl
6	51	14.2	177	11	US-09-791-932-95	Sequence 95, Appl
7	11	3.1	43	12	US-10-124-557-12	Sequence 12, Appl
8	11	3.1	94	10	US-09-864-864-1273	Sequence 1273, Ap
9	11	3.1	104	9	US-09-864-761-41522	Sequence 41522, A
10	11	3.1	112	10	US-09-738-626-5042	Sequence 5042, App
11	11	3.1	134	10	US-09-738-626-5042	Sequence 812, App
12	11	3.1	187	12	US-10-148-687-4	Sequence 4, Appl
13	11	3.1	192	14	US-10-124-557-90	Sequence 90, Appl
14	11	3.1	199	10	US-09-764-864-1291	Sequence 1291, Ap
15	11	3.1	201	10	US-09-764-864-836	Sequence 836, App
16	11	3.1	204	14	US-10-124-557-92	Sequence 92, Appl
17	11	3.1	209	14	US-10-124-557-132	Sequence 132, Appl
18	11	3.1	217	14	US-10-124-557-94	Sequence 94, Appl
19	11	3.1	220	14	US-10-124-557-76	Sequence 76, Appl
20	11	3.1	231	14	US-10-124-557-96	Sequence 96, Appl
21	11	3.1	243	14	US-10-124-557-30	Sequence 30, Appl
22	11	3.1	243	11	US-09-533-029-34	Sequence 34, Appl
23	11	3.1	243	11	US-09-934-455-416	Sequence 416, Ap
24	11	3.1	243	12	US-10-374-780A-2074	Sequence 2074, Ap
25	11	3.1	275	12	US-10-148-68?-3	Sequence 3, Appl
26	11	3.1	296	14	US-10-124-557-70	Sequence 70, Appl
27	11	3.1	310	15	US-10-007-521-22	Sequence 22, Appl
28	11	3.1	324	9	US-09-734-417-2	Sequence 2, Appl
29	11	3.1	328	12	US-10-148-687-1	Sequence 1, Appl
30	11	3.1	372	14	US-10-124-557-64	Sequence 64, Appl
31	11	3.1	403	12	US-10-369-493-4127	Sequence 4127, Ap
32	11	3.1	422	14	US-10-124-557-68	Sequence 68, Appl
33	11	3.1	423	14	US-10-124-557-66	Sequence 66, Appl
34	11	3.1	447	11	US-09-833-245-1720	Sequence 1720, Ap
35	11	3.1	448	11	US-09-386-32421	Sequence 32421, A
36	11	3.1	448	11	US-10-108-260A-2481	Sequence 2481, Ap
37	11	3.1	450	11	US-10-108-260A-2481	Sequence 268, App
38	11	3.1	450	15	US-10-153-668-380	Sequence 380, App
39	11	3.1	450	15	US-10-153-668-382	Sequence 382, App
40	11	3.1	463	14	US-10-124-557-54	Sequence 54, Appl
41	11	3.1	503	10	US-09-801-368-424	Sequence 424, App
42	11	3.1	503	12	US-10-369-493-1943	Sequence 1943, Ap
43	11	3.1	506	12	US-10-369-493-4115	Sequence 4115, Ap
44	11	3.1	544	12	US-10-032-585-7924	Sequence 7924, Ap
45	11	3.1	547	12	US-10-264-049-2651	Sequence 2651, Ap

```
                                                ALIGNMENTS

RESULT 1
US-10-041-615-34
; Sequence 34, Application US/10041615
; Publication No.US2004001403BA1
; GENERAL INFORMATION:
;  APPLICANT: Casman, Stacie J
;  APPLICANT: Edinger, Shlomit R
;  APPLICANT: Ellerman, Karen
;  APPLICANT: Smithson, Glennda
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT: Padigaru, Muralidhara
;  TITLE OF INVENTION: No. US2004001403BA1e1 GPCR-Like Proteins and Nucleic Acids Encodi
;  FILE REFERENCE: 21402-233-061
;  CURRENT APPLICATION NUMBER: US/10/041,615
;  CURRENT FILING DATE: 2003-01-29
;  PRIOR APPLICATION NUMBER: 60/259,552
;  PRIOR FILING DATE: 2001-01-03
;  PRIOR APPLICATION NUMBER: 60/260,544
;  PRIOR FILING DATE: 2001-01-09
;  PRIOR APPLICATION NUMBER: 60/277,405
;  PRIOR FILING DATE: 2001-03-20
;  NUMBER OF SEQ ID NOS: 174
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 34
;  LENGTH: 360
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-041-615-34

Query Match             100.0%;  Score 360;  DB 12;  Length 360;
Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  360;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MIKLGPYFTPPTKIKTKIKDLNVESQTVKIFLRSLFYTALLVFSALGNILALCLTCQKSR 60

Qy         61 KINCTGIYLVHLAVSDLLFTVALPGRVVCYVLGSSWPFGKGLCRLTAFVLYTDTYGGVYL 120
```

```
Db    61  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          KINCTGYLVLHLAVSDLLFTVALPGRVVCVYLGSSWPFGKGLCRLTAFVLFLYTDTYGGVYL  120

Qy   121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  MACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLAC  180

Qy   181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  MEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTAREDPVTSRKGRHWR  240

Qy   241  GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLFLPSCAERRAFLLSLQATVALMNMNCGI  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GCLLTLMLVAVVVCFSPYHLNIKQFMARGMLFLPSCAERRAFLLSLQATVALMNMNCGI  300

Qy   301  TPIIYFFASTHYRKNLGILKLKGSSSSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  TPIIYFFASTHYRKNLGILKLKGSSSSSSSSSSSSSSSTPGKASSETPSITQARGSMFLAEHVV  360
```

RESULT 2
US-10-041-615-109
; Sequence 109, Application US20040014038A1
; Publication No. US20040014038A1
; GENERAL INFORMATION:
; APPLICANT: Casman, Stacie J
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glennda
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014038A1el GPCR-Like Proteins and Nucleic Acids Encodi
; FILE REFERENCE: 21402-233-061
; CURRENT APPLICATION NUMBER: US/10/041,615
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: 60/259,552
; PRIOR FILING DATE: 2001-01-03
; PRIOR APPLICATION NUMBER: 60/260,544
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/277,405
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 109
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-615-109

Query Match   30.6%; Score 110; DB 12; Length 229;
Best Local Similarity 100.0%; Pred. No. 5.9e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   117  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  176
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   116  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  175

Qy   177  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  226
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   176  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  225
```

RESULT 3
US-10-017-161-1620
; Sequence 1620, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152

; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1620
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1620

Query Match   30.6%; Score 110; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   117  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  176
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    58  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  117

Qy   177  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  226
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   118  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  167
```

RESULT 4
US-10-292-798-1294
; Sequence 1294, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1294
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1294

Query Match   30.6%; Score 110; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   117  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  176
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    58  GVYLMACVSVDHYPAVVCAHWGPRLRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVG  117

Qy   177  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  226
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   118  KLACMEYSSMESVLGLPLMVLVAFAIGFCGPVGIILSCYMKITWKLCSTA  167
```

RESULT 5
US-09-801-944B-198
; Sequence 198, Application US/09801944B
; Publication No. US20040014169A1
; GENERAL INFORMATION:
; APPLICANT: Vogel1, Gabriel
; APPLICANT: Wood, Linda S.
; TITLE OF INVENTION: No. US20040014169A1el G Protein-Coupled Receptors
; FILE REFERENCE: 0010US1
; CURRENT APPLICATION NUMBER: US/09/801,944B

```
;   CURRENT FILING DATE: 2003-08-01
;   PRIOR APPLICATION NUMBER: 60/187,828
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,715
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,929
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,930
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,825
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,833
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,830
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,829
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,582
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/187,581
;   PRIOR FILING DATE: 2000-03-08
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 273
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 198
;   LENGTH: 177
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-801-944B-198

Query Match      19.4%;  Score 70;  DB 12;  Length 177;
Best Local Similarity  100.0%;  Pred. No. 4.1e-58;
Matches    70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy  233 SRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVA 292
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 SRKGRHWRGCLLTLMLVAVVVCFSPYHLNIKQFMARGMLHLPSCAERRAFLLSLQATVA 60

Qy  293 LMMNNCGITP 302
        ||||||||||
Db   61 LMMNNCGITP 70

RESULT 6
US-09-791-932-95
; Sequence 95, Application US/09791932
; Publication No. US20030003451A1
; GENERAL INFORMATION:
;   APPLICANT: Vogeli, Gabriel
;   APPLICANT: Parodi, Luis A.
;   APPLICANT: Hiebsch, Ronald R.
;   APPLICANT: Lind, Peter
;   APPLICANT: Kaytes, Paul S.
;   APPLICANT: Ruff, Valerie
;   APPLICANT: Huff, Rita M.
;   APPLICANT: Wood, Linda S.
;   TITLE OF INVENTION:  No. US20030003451A1e1  G Protein-Coupled Receptors Cross-Referen
;   FILE REFERENCE:  00325.US1
;   CURRENT APPLICATION NUMBER: US/09/791,932
;   CURRENT FILING DATE: 2001-02-23
;   PRIOR APPLICATION NUMBER: 60/184,305
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,304
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,303
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,397
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/184,247
;   PRIOR FILING DATE: 2000-02-23
;   PRIOR APPLICATION NUMBER: 60/188,880
;   PRIOR FILING DATE: 2000-03-13
;   PRIOR APPLICATION NUMBER: 60/217,369
```

```
;   PRIOR FILING DATE: 2000-07-11
;   PRIOR APPLICATION NUMBER: 60/217,370
;   PRIOR FILING DATE: 2000-07-11
;   PRIOR APPLICATION NUMBER: 60/218,492
;   PRIOR FILING DATE: 2000-07-20
;   PRIOR APPLICATION NUMBER: 60/186,810
;   PRIOR FILING DATE: 2000-03-03
;   PRIOR APPLICATION NUMBER: 60/188,064
;   PRIOR FILING DATE: 2000-03-09
;   PRIOR APPLICATION NUMBER: 60/186,457
;   PRIOR FILING DATE: 2000-03-02
;   PRIOR APPLICATION NUMBER: 60/213,861
;   PRIOR FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: 60/194,344
;   PRIOR FILING DATE: 2000-04-03
;   PRIOR APPLICATION NUMBER: 60/218,337
;   PRIOR FILING DATE: 2000-07-14
;   NUMBER OF SEQ ID NOS: 184
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 95
;   LENGTH: 177
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-791-932-95

Query Match      14.2%;  Score 51;  DB 11;  Length 177;
Best Local Similarity  100.0%;  Pred. No. 4.3e-40;
Matches    51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy  141 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLG 191
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 LRTAGRARLVCVAIWTLVLLQTMPLLLMPMTKPLVGKLACMEYSSMESVLG 98

RESULT 7
US-10-124-557-12
; Sequence 12, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
;   APPLICANT: Turner, Katherine
;             Clark, Stephen C.
;             Jacobs, Kenneth
;             Hewick, Rodney M.
;             Gesner, Thomas G.
;   TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;   NUMBER OF SEQUENCES: 143
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genetics Institute, Inc.
;     STREET: 87 CambridgePark Drive
;     CITY: Cambridge
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 02140
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/124,557
;     FILING DATE: 16-Apr-2002
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/643,502
;     FILING DATE: 18-JAN-1991
;     APPLICATION NUMBER: US 07/546,114
;     FILING DATE: 29-JUN-1990
;     APPLICATION NUMBER: US 07/457,196
;     FILING DATE: 29-DEC-1989
;     APPLICATION NUMBER: US 07/390,901
;     FILING DATE: 08-AUG-1989
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Cserr, Luann
```

```
          REGISTRATION NUMBER: 31,822
          REFERENCE/DOCKET NUMBER: GI 5190
          TELECOMMUNICATION INFORMATION:
          TELEPHONE: (617)876-1170
          TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 12:
          SEQUENCE CHARACTERISTICS:
          LENGTH: 43 amino acids
          TYPE: amino acid
          TOPOLOGY: linear
          MOLECULE TYPE: protein
          SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-124-557-12

Query Match     3.1%;  Score 11;  DB 14;  Length 43;
Best Local Similarity  100.0%;  Pred. No. 0.011;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          ||||||||||
Db     12 SSSSSSSSST 22

RESULT 8
US-09-764-864-1273
; Sequence 1273, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1273
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1273

Query Match     3.1%;  Score 11;  DB 10;  Length 94;
Best Local Similarity  100.0%;  Pred. No. 0.021;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    324 GSSSSSSSSS 334
          ||||||||||
Db     17 GSSSSSSSSS 27

RESULT 9
US-09-864-761-41522
; Sequence 41522, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41522
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL161747.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
; OTHER INFORMATION: EST HUMAN HIT: BF340131.1, EVALUE 2.00e-24
; OTHER INFORMATION: SWISSPROT HIT: P53235, EVALUE 1.50e+00
US-09-764-761-41522

Query Match     3.1%;  Score 11;  DB 9;  Length 104;
Best Local Similarity  100.0%;  Pred. No. 0.023;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          ||||||||||
Db     83 SSSSSSSSST 93

RESULT 10
US-09-764-864-812
; Sequence 812, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 812
; LENGTH: 112
; TYPE: PRT
```

; ORGANISM: Homo sapiens
US-09-764-864-812

Query Match   3.1%; Score 11; DB 10; Length 112;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   324 GSSSSSSSSSS 334
          |||||||||||
Db   35 GSSSSSSSSSS 45

RESULT 11
US-09-738-626-5042
; Sequence 5042, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5042
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5042

Query Match   3.1%; Score 11; DB 10; Length 134;
Best Local Similarity 100.0%; Pred. No. 0.029;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   325 SSSSSSSSST 335
          ||||||||||
Db   42 SSSSSSSSST 52

RESULT 12
US-10-148-687-4
; Sequence 4, Application US/10148687
; Publication No. US20030185836A1
; GENERAL INFORMATION:
; APPLICANT: WINTER, Gerhard
; APPLICANT: SLADE, Martin Basil
; APPLICANT: WILLIAMS, Keith Leslie
; APPLICANT: GOOLEY, Andrew Arthur
; APPLICANT: Macquarie Research Ltd
; TITLE OF INVENTION: Cryptosporidium sporozoite antigens
; FILE REFERENCE: 047763-5019-US
; CURRENT APPLICATION NUMBER: US/10/148,687
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: PCT/AU00/01492
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: AU PQ4400
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 67

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-10-148-687-4

Query Match   3.1%; Score 11; DB 12; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   324 GSSSSSSSSSS 334
          |||||||||||
Db   6 GSSSSSSSSSS 16

RESULT 13
US-10-124-557-90
; Sequence 90, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;   Clark, Stephen C.
;   Jacobs, Kenneth
;   Hewick, Rodney M.
;   Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genetics Institute, Inc.
;   STREET: 87 CambridgePark Drive
;   CITY: Cambridge
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02140
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/124,557
;   FILING DATE: 16-Apr-2002
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/643,502
;   FILING DATE: 18-JAN-1991
;   APPLICATION NUMBER: US 07/546,114
;   FILING DATE: 29-JUN-1990
;   APPLICATION NUMBER: US 07/457,196
;   FILING DATE: 29-DEC-1989
;   APPLICATION NUMBER: US 07/390,901
;   FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
;   NAME: Cserr, Luann
;   REGISTRATION NUMBER: 31,822
;   REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)876-1170
;   TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 192 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 90:
US-10-124-557-90

Query Match   3.1%; Score 11; DB 14; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      325 SSSSSSSSSST 335
            ||||||||||||
Db      168 SSSSSSSSSST 178

RESULT 14
US-09-764-864-1291
; Sequence 1291, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1291
;   LENGTH: 199
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (187)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1291

Query Match          3.1%;    Score 11;   DB 10;   Length 199;
Best Local Similarity   100.0%;   Pred. No. 0.041;
Matches  11;   Conservative   0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy      324 GSSSSSSSSS 334
            ||||||||||
Db       31 GSSSSSSSSS 41

RESULT 15
US-09-764-864-836
; Sequence 836, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 836
;   LENGTH: 201
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-764-864-836

Query Match          3.1%;    Score 11;   DB 10;   Length 201;
Best Local Similarity   100.0%;   Pred. No. 0.042;
Matches  11;   Conservative   0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy      324 GSSSSSSSSS 334
            ||||||||||
Db       33 GSSSSSSSSS 43

Search completed: February 19, 2004, 19:40:41
Job time : 41 secs
```

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 19, 2004, 19:35:36 ; Search time 20 Seconds
                        (without alignments)
                        1731.036 Million cell updates/sec

Title:     US-10-041-615-34
Perfect score: 360
Sequence:  1 MIKLGPYFTPPTKIKTKIKD........SETPSITQARGSMFLAEHVV 360

Scoring table: OLIGO
           Gapop 60.0 , Gapext 60.0

Searched:  283308 seqs, 96168682 residues

Word size : 10

Total number of hits satisfying chosen parameters:      218

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  PIR 76:*
            1:  pir1:*
            2:  pir2:*
            3:  pir3:*
            4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	3.1	135	2	A42098	transcription fact
2	11	3.1	162	2	C85356	glycine-rich prote
3	11	3.1	192	2	S59537	heat shock transcr
4	11	3.1	243	2	T51989	ethylene responsiv
5	11	3.1	262	2	T40941	hypothetical serin
6	11	3.1	284	2	A35419	neutrophil protein
7	11	3.1	288	2	B40722	homeotic protein 1
8	11	3.1	306	2	T47832	hypothetical prote
9	11	3.1	323	2	G96544	hypothetical prote
10	11	3.1	416	2	G83656	hypothetical prote
11	11	3.1	446	2	T45525	WSC4 homolog [impo
12	11	3.1	447	2	E83465	conserved hypothet
13	11	3.1	467	2	S30839	UTR2 protein - yea
14	11	3.1	503	2	S63257	probable membrane
15	11	3.1	568	2	I39411	AF-9 protein - hum
16	11	3.1	583	1	S22544	transcription fact
17	11	3.1	641	1	A41932	fibrinogen alpha-I
18	11	3.1	710	2	S28014	outD protein - Erw
19	11	3.1	734	2	B42680	nucleolus-cytoplas
20	11	3.1	741	2	I48694	probable transcrip
21	11	3.1	742	2	A49672	transcription fact
22	11	3.1	772	2	A55004	transcription fact
23	11	3.1	779	2	T49717	related to BCS1 pr
24	11	3.1	962	2	S03818	carboxymethylcellu
25	11	3.1	1019	2	T30148	hypothetical prote
26	11	3.1	1052	2	C64221	hypothetical 114K
27	11	3.1	1215	2	I52882	autoantigen - huma
28	11	3.1	1237	2	A34598	ecdysone-induced p
29	11	3.1	1341	2	S50366	probable membrane
30	11	3.1	1390	2	T14004	trfA protein - sli
31	11	3.1	1394	2	B34598	ecdysone-induced p
32	11	3.1	1443	2	S05979	steroid hormone re
33	11	3.1	1723	2	H86557	polymorphic membra
34	11	3.1	1723	2	E72067	polymorphic membra
35	11	3.1	1732	2	C81601	polymorphic membra
36	11	3.1	1823	2	S28974	vitellogenin precu
37	11	3.1	1912	2	T29088	vitellogenin I pre
38	10	2.8	98	2	S13623	ht-en protein - le
39	10	2.8	105	2	C86242	protein T16B5.1 [i
40	10	2.8	107	2	B85356	glycine-rich prote
41	10	2.8	109	2	T12208	probable cytochrom
42	10	2.8	125	2	S69871	hypothetical prote
43	10	2.8	135	2	T09876	dehydrin - upland
44	10	2.8	145	2	S04042	embryonic abundant
45	10	2.8	149	2	T23179	hypothetical prote

ALIGNMENTS

```
RESULT 1
A42098
transcription factor Oct-2B, octamer-binding - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C;Accession: A42098
R;Tanaka, M.; Lai, J.S.; Herr, W.
Cell 68, 755-767, 1992
A;Title: Promoter-selective activation domains in Oct-1 and Oct-2 direct differential ac
A;Reference number: A42098; MUID:92154688; PMID:1739980
A;Note: nucleotide sequence is not given
A;Accession: A42098
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-135 <TAN>
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match           3.1%;  Score 11;  DB 2;  Length 135;
Best Local Similarity 100.0%;  Pred. No. 0.0042;
Matches 11;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    325 SSSSSSSSSST 335

Db    103 SSSSSSSSSST 113


RESULT 2
C85356
glycine-rich protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C;Accession: C85356
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: C85356
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-162 <STO>
A;Cross-references: GB:NC_001268; NID:g7269947; PIDN:CAB79764.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g30460
A;Map position: 4
C;Superfamily: glycine-rich cell wall structural protein 1

Query Match           3.1%;  Score 11;  DB 2;  Length 162;
Best Local Similarity 100.0%;  Pred. No. 0.005;
Matches 11;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    324 GSSSSSSSSSS 334
```

```
Db     69 GSSSSSSSSS 79
          ||||||||||

RESULT 3
S59537
heat shock transcription factor HSF21 - soybean (fragment)
C;Species: Glycine max (soybean)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 02-Sep-2000
C;Accession: S59537; S52303
R;Czarnecka-Verner, E.; Yuan, C.X.; Fox, P.C.; Gurley, W.B.
Plant Mol. Biol. 29, 37-51, 1995
A;Title: Isolation and characterization of six heat shock transcription factor cDNA clon
A;Reference number: S59537; MUID:96017612; PMID:7559166
A;Accession: S59537
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-192 <CZA>
A;Cross-references: EMBL:Z46952
C;Genetics:
A;Gene: HSF21
C;Superfamily: tomato heat shock transcription factor HSF24; HSF DNA-binding domain homo
C;Keywords: DNA binding; heat shock; leucine zipper; nucleus; stress-induced protein; tr
F;46-141/Domain: HSF DNA-binding domain homology <HSF>

Query Match          3.1%;  Score 11;  DB 2;  Length 192;
Best Local Similarity   100.0%;  Pred. No. 0.0058;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    324 GSSSSSSSSS 334
          ||||||||||
Db     19 GSSSSSSSSS 29

RESULT 4
T51989
ethylene responsive element binding factor 2 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T51989
R;Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M.
Plant Cell 12, 393-404, 2000
A;Title: Arabidopsis ethylene responsive element binding factors act as transcriptional
A;Reference number: 225893
A;Accession: T51989
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-243 <FUJ>
A;Cross-references: EMBL:AB008104; PIDN:BAA32419.1
C;Genetics:
A;Gene: ERF-2

Query Match          3.1%;  Score 11;  DB 2;  Length 243;
Best Local Similarity   100.0%;  Pred. No. 0.0072;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          ||||||||||
Db    193 SSSSSSSSST 203

RESULT 5
T40941
hypothetical serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: T40941
R;Lucas, M.; Gaillardin, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A;Reference number: 221959
A;Accession: T40941
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-262 <LUC>
A;Cross-references: EMBL:AL035259; PIDN:CAA22863.1; GSPDB:GN00068; SPDB:SPCC1322.10
A;Experimental source: strain 972h-; cosmid c1322
C;Genetics:
A;Gene: SPDB:SPCC1322.10
A;Map position: 3
C;Superfamily: serine-rich protein

Query Match          3.1%;  Score 11;  DB 2;  Length 262;
Best Local Similarity   100.0%;  Pred. No. 0.0077;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    326 SSSSSSSSTP 336
          ||||||||||
Db    156 SSSSSSSSTP 166

RESULT 6
A35419
neutrophil protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 05-Nov-1999
C;Accession: A35419
R;Bellavite, P.; Bazzoni, F.; Cassatella, M.A.; Hunter, K.J.; Bannister, J.V.
Biochem. Biophys. Res. Commun. 170, 915-922; 1990
A;Title: Isolation and characterization of a cDNA clone for a novel serine-rich neutroph
A;Reference number: A35419; MUID:90343818; PMID:1696479
A;Accession: A35419
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-284 <BEL>
A;Cross-references: GB:M55701; NID:g164672; PIDN:AAA63449.1; PID:g164673

Query Match          3.1%;  Score 11;  DB 2;  Length 284;
Best Local Similarity   100.0%;  Pred. No. 0.0082;
Matches  11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325 SSSSSSSSST 335
          ||||||||||
Db    145 SSSSSSSSST 155

RESULT 7
B40722
homeotic protein lin-32 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Nov-2000
C;Accession: B40722; A40720
R;Wang, B.B.; Mueller-Immergluck, M.M.; Austin, J.; Robinson, N.T.; Chisholm, A.; Kenyor
Cell 74, 29-42, 1993
A;Title: A homeotic gene cluster patterns the anteroposterior body axis of Caenorhabditi
A;Reference number: A40722; MUID:93327429; PMID:8101474
A;Accession: B40722
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-288 <WAN>
A;Cross-references: GB:L19248
R;Clark, S.G.; Chisholm, A.D.; Horvitz, H.R.
Cell 74, 43-55, 1993
A;Title: Control of cell fates in the central body region of Caenorhabditis elegans by t
A;Reference number: A40720; MUID:93327430; PMID:8101475
A;Accession: A40720
A;Status: preliminary
A;Molecule type: mRNA
A;Gene: lin-32
C;Genetics:
A;Gene: lin-32
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;201-257/Domain: homeobox homology <HOX>

Query Match          3.1%;  Score 11;  DB 2;  Length 288;
```

Best Local Similarity   100.0%;   Pred. No. 0.0083;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  325 SSSSSSSSST 335
          |||||||||
Db   59 SSSSSSSSST  69
```

RESULT 8
T47832
hypothetical protein T2O9.60 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C;Accession: T47832
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24475
A;Accession: T47832
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-306 <NYA>
A;Cross-references: EMBL:AL138658
A;Experimental source: cultivar Columbia; BAC clone T2O9
C;Genetics:
A;Map position: 3
A;Note: T2O9.60
C;Superfamily: RING finger homology
F;165-215/Domain: RING finger homology <RRN>

Query Match   3.1%;   Score 11;   DB 2;   Length 306;
Best Local Similarity   100.0%;   Pred. No. 0.0088;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  324 SSSSSSSSSS 334
          ||||||||||
Db  123 SSSSSSSSSS 133
```

RESULT 9
G96544
hypothetical protein F8A12.3 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G96544
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.;
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, B.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:22016719; PMID:11130712
A;Accession: G96544
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <STO>
A;Cross-references: GB:AE005173; GB:AE005173; NID:gi11094707; PIDN:AAG29643.1; GSPDB:GN00141
C;Genetics:
A;Gene: F8A12.3
A;Map position: 1

Query Match   3.1%;   Score 11;   DB 2;   Length 323;
Best Local Similarity   100.0%;   Pred. No. 0.0093;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  324 SSSSSSSSSS 334
          ||||||||||
Db  165 SSSSSSSSSS 175
```

RESULT 10
G83656
hypothetical protein BH0055 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: G83656
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G83656
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-416 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:gi10172612; PIDN:BAB03774.1; GSPDB:GN0C
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0055

Query Match   3.1%;   Score 11;   DB 2;   Length 416;
Best Local Similarity   100.0%;   Pred. No. 0.012;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  324 GSSSSSSSSSS 334
          |||||||||||
Db  285 GSSSSSSSSSS 295
```

RESULT 11
T45525
WSC4 homolog [imported] - yeast (Kluyveromyces marxianus var. lactis)
C;Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 31-Jan-2000
C;Accession: T45525
R;Bao, W.G.; Fukuhara, H.
submitted to the EMBL Data Library, July 1999
A;Description: The ubiquitin-encoding genes of Kluyveromyces lactis.
A;Reference number: 223000
A;Accession: T45525
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-446 <BAO>
A;Cross-references: EMBL:AJ243800; PIDN:CAB50897.1
A;Experimental source: strain 2359/152
C;Genetics:
A;Gene: wsc4

Query Match   3.1%;   Score 11;   DB 2;   Length 446;
Best Local Similarity   100.0%;   Pred. No. 0.012;
Matches   11;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  325 SSSSSSSSST 335
          ||||||||||
Db  172 SSSSSSSSST 182
```

RESULT 12
E83465
conserved hypothetical protein PA1451 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83465
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83465
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <STO>
A;Cross-references: GB:AE004574; GB:AE004091; NID:gi9947391; PIDN:AAG04840.1; GSPDB:GN001

A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1451

Query Match   3.1%; Score 11; DB 2; Length 447;
Best Local Similarity   100.0%; Pred. No. 0.012;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
           ||||||||||
Db   417 GSSSSSSSSS 427

RESULT 13
S30839
UTR2 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YEL040w
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 06-Feb-1998
R;Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A;Reference number: S30812
A;Accession: S30839
A;Molecule type: DNA
A;Residues: 1-467 <MUL>
A;Cross-references: GB:U18779; EMBL:L10830; NID:g603625; PID:g603639
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A;Description: The sequence of S. cerevisiae cosmids 8199, 8334, and 9871.
A;Reference number: S50491
A;Accession: S50504
A;Molecule type: DNA
A;Residues: 1-467 <DIE>
A;Cross-references: EMBL:U18779; NID:g603625; PID:g603639; MIPS:YEL040w
R;Melnick, L.; Sherman, F.
J. Mol. Biol. 233, 372-388, 1993
A;Title: The gene clusters ARC and COR on chromosomes 5 and 10, respectively, of Sacchar
A;Reference number: S38543; MUID:94016558; PMID:8411151
A;Accession: S38545
A;Molecule type: DNA
A;Residues: 121-129,'V',131-290,'R',292-353,'C',355-467 <MEL>
A;Cross-references: EMBL:S66130; NID:g430829; PID:g430830
C;Genetics:
A;Gene: SGD:UTR2
A;Cross-references: SGD:S0000766; MIPS:YEL040w
A;Map position: 5L

Query Match   3.1%; Score 11; DB 2; Length 467;
Best Local Similarity   100.0%; Pred. No. 0.013;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSST 335
           ||||||||||
Db   394 SSSSSSSSST 404

RESULT 14
S63257
probable membrane protein YNL283c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N0583
C;Species: Saccharomyces cerevisiae
C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 06-Feb-1998
R;Messenguy, F.; Dubois, E.; Vierendeels, F.; Scherens, B.; Pierard, A.; Glansdorff, N.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S63245
A;Accession: S63257
A;Molecule type: DNA
A;Residues: 1-503 <MES>
A;Cross-references: EMBL:Z71559; NID:g1302356; PID:g1302357; MIPS:YNL283c
A;Experimental source: strain S288C
C;Genetics:

A;Gene: SGD:WSC2
A;Cross-references: SGD:S0005227; MIPS:YNL283c
A;Map position: 14L
C;Keywords: transmembrane protein
F;330-346/Domain: transmembrane #status predicted <TMM>

Query Match   3.1%; Score 11; DB 2; Length 503;
Best Local Similarity   100.0%; Pred. No. 0.014;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSST 335
           ||||||||||
Db   183 SSSSSSSSST 193

RESULT 15
I39411
AF-9 protein - human
C;Species: Homo sapiens (man)
C;Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997
C;Accession: I39411
R;Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Gale, R.P.; Lang
Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993
A;Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leuk
A;Reference number: A47440; MUID:93281633; PMID:8506309
A;Accession: I39411
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-568 <NAK>
A;Cross-references: GB:L13744; NID:g306448; PID:g306449
C;Genetics:
A;Gene: GDB:MLLT3; AF-9
A;Cross-references: GDB:138172; OMIM:159558
A;Map position: 9p22-9p22

Query Match   3.1%; Score 11; DB 2; Length 568;
Best Local Similarity   100.0%; Pred. No. 0.015;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSST 335
           ||||||||||
Db   181 SSSSSSSSST 191

Search completed: February 19, 2004, 19:39:16
Job time : 23 secs

```
              GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   February 19, 2004, 19:34:31 ; Search time 17 Seconds
                                (without alignments)
                       995.860 Million cell updates/sec

Title:          US-10-041-615-34
Perfect score:  360
Sequence:       1 MIKLGPYFTPPTKITKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       127863 seqs, 47026705 residues

Word size :     10

Total number of hits satisfying chosen parameters:   112

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      SwissProt_41:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	11	3.1	243	1	ERF2_ARATH	O80338	arabidopsis
2	11	3.1	253	1	LI39_CAEEL	P34684	caenorhabdi
3	11	3.1	347	1	UTR2_YEAST	P32623	saccharomyc
4	11	3.1	503	1	WSC2_YEAST	P53832	saccharomyc
5	11	3.1	546	1	PHR2_CANAL	O13318	candida alb
6	11	3.1	568	1	AF9_HUMAN	P42568	homo sapien
7	11	3.1	624	1	RN12_HUMAN	Q9NVW2	homo sapien
8	11	3.1	634	1	HWP1_CANAL	P46593	candida alb
9	11	3.1	641	1	FIB2_PETNA	P33573	petromyzon
10	11	3.1	704	1	NP14_RAT	P41777	rattus norv
11	11	3.1	706	1	CT67_MOUSE	P59114	mus musculu
12	11	3.1	710	1	GSQD_ERWCH	Q01565	erwinia chr
13	11	3.1	741	1	NFL1_MOUSE	Q61985	mus musculu
14	11	3.1	772	1	NFL1_HUMAN	Q14494	h nuclear f
15	11	3.1	962	1	GUNA_PSEFL	P10476	pseudomonas
16	11	3.1	1052	1	MGPC_MYCGE	P22747	mycoplasma
17	11	3.1	1237	1	E75A_DROME	P17671	drosophila
18	11	3.1	1341	1	YL78_YEAST	Q05854	saccharomyc
19	11	3.1	1394	1	E75B_DROME	P17672	drosophila
20	11	3.1	1443	1	E75C_DROME	P13055	drosophila
21	11	3.1	1723	1	PM2C_CHLPN	Q9Z812	chlamydia p
22	11	3.1	1823	1	VIT_ICHUN	Q91062	ichthyomyzo
23	11	3.1	1912	1	VIT1_CHICK	P87498	gallus gall
24	11	3.1	2004	1	CHDB_HUMAN	Q9HCK8	homo sapien
25	10	3.1	98	1	HMEN_HELTR	P23397	helobdella
26	10	2.8	145	1	DH1_GOSHI	P09442	gossypium h
27	10	2.8	170	1	HSB7_HUMAN	Q9UBY9	homo sapien
28	10	2.8	215	1	FGFA_RAT	P70492	rattus norv
29	10	2.8	220	1	SER2_GALME	O96615	galleria me
30	10	2.8	260	1	OLG1_MOUSE	Q9JKN5	mus musculu
31	10	2.8	304	1	WR13_ARATH	Q9SVB7	arabidopsis
32	10	2.8	334	1	LYTE_BACSU	P54421	bacillus su
33	10	2.8	336	1	CAHC_ARATH	P27140	arabidopsis
34	10	2.8	338	1	OTX3_BRARE	Q90267	brachydanio
35	10	2.8	344	1	YB1E_SCHPO	P87179	schizosacch
36	10	2.8	347	1	VIT3_CHICK	Q91025	gallus gall
37	10	2.8	352	1	STSY_CATRO	P18417	catharanthu
38	10	2.8	364	1	NK61_MESAU	Q60554	mesocricetu
39	10	2.8	365	1	NK61_MOUSE	Q99MA9	mus musculu
40	10	2.8	365	1	NK61_RAT	O35762	rattus norv
41	10	2.8	367	1	NK61_HUMAN	P78426	homo sapien
42	10	2.8	398	1	CAR4_RHINI	Q03700	rhizopus ni
43	10	2.8	399	1	HM39_CAEEL	Q22812	caenorhabdi
44	10	2.8	406	1	SR40_YEAST	P32583	saccharomyc
45	10	2.8	413	1	E2F4_HUMAN	Q16254	homo sapien

ALIGNMENTS

```
RESULT 1
ERF2_ARATH
ID   ERF2_ARATH     STANDARD;      PRT;    243 AA.
AC   O80338;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ethylene responsive element binding factor 2 (AtERF2).
GN   ERF2 OR ERF-2 OR AT5G47220 OR MQL5.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20181733; PubMed=10715325;
RA   Fujimoto S.Y., Ohta M., Usui A., Shinshi H., Ohme-Takagi M.;
RT   "Arabidopsis ethylene responsive element binding factors act as
RT   transcriptional activators or repressors of GCC box mediated gene
RT   expression.";
RL   Plant Cell 12:393-404(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
     -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO THE GCC-BOX
CC   -!- PATHOGENESIS-RELATED PROMOTER ELEMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 AP2/ERF domain.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on its
CC   use  by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed.  Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AB008104; BAA32419.1; -.
DR   EMBL; AB018117; BAA97155.1; -.
DR   PIR; T51989; T51989.
DR   HSSP; O80037; 2GCC.
DR   TRANSFAC; T04634; -.
DR   InterPro; IPR001471; TF_ERF.
DR   Pfam; PF00847; AP2-domain; 1.
DR   ProDom; PD001423; TF_ERF; 1.
DR   SMART; SM00380; AP2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein.
FT   DOMAIN       23     26       POLY-GLY.
```

```
FT   DNA_BIND    114     179
FT   DOMAIN      192     206     AP2/ERF.
SQ   SEQUENCE   243 AA;  26797 MW;  POLY-SER.
                                   C9A4C9791249B5D4 CRC64;

Query Match             3.1%;  Score 11;  DB 1;  Length 243;
Best Local Similarity  100.0%;  Pred. No. 0.0045;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325  SSSSSSSSSST  335
           |||||||||||
Db    193  SSSSSSSSSST  203

RESULT 2
L139_CAEEL
ID   L139_CAEEL    STANDARD;  PRT;  253 AA.
AC   P34684;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein lin-39.
GN   LIN-39 OR CEH-15 OR CO7H6.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=9327430; PubMed=8101475;
RA   Clark S.G., Chisholm A.D., Horvitz H.R.;
RT   "Control of cell fates in the central body region of C. elegans by
RT   the homeobox gene lin-39.";
RL   Cell 74:43-55(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9327429; PubMed=8101474;
RA   Wang B.B., Mueller-Immergluck M.M., Austin J., Robinson N.T.,
RA   Chisholm A.D., Kenyon C.;
RT   "A homeotic gene cluster patterns the anteroposterior body axis of C.
RT   elegans.";
RL   Cell 74:29-42(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Macri C., Vaudin M.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CONTROLS THE MIGRATION OF NEUROBLASTS AND THE
CC       SUBSEQUENT MID-BODY REGION-SPECIFIC DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
CC       "DEFORMED" SUBFAMILY.
```

```
DR   EMBL; L19639; AAC37168.1; -.
DR   EMBL; L19248; AAB04137.1; -.
DR   EMBL; AC006605; AAK55445.1; -.
DR   PIR; B40722; B40722.
DR   HSSP; P02833; 1SAN.
DR   WormPep; CO7H6.7; CE03975.
DR   TRANSFAC; T03369; -.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00025; ANTENNAPEDIA.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PROSITE; PS00032; ANTENNAPEDIA; 1.
KW   Homeobox; DNA-binding; Nuclear protein; Developmental protein.
FT   DOMAIN      20      36      POLY-SER.
FT   DOMAIN     121     129      POLY-ASP.
FT   SITE       148     153      ANTP-TYPE HEXAPEPTIDE.
FT   DNA_BIND   165     224      HOMEOBOX.
FT   DOMAIN     228     248      PRO-RICH.
SQ   SEQUENCE   253 AA;  27814 MW;  6B59DE9568B834A1 CRC64;

Query Match             3.1%;  Score 11;  DB 1;  Length 253;
Best Local Similarity  100.0%;  Pred. No. 0.0047;
Matches   11;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    325  SSSSSSSSSST  335
           |||||||||||
Db     24  SSSSSSSSSST  34

RESULT 3
UTR2_YEAST
ID   UTR2_YEAST    STANDARD;  PRT;  347 AA.
AC   P32623;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   UTR2 protein (Unknown transcript 2 protein).
GN   UTR2 OR YEL040W OR SYGP-ORF18.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B-6441;
RX   MEDLINE=94016558; PubMed=8411151;
RA   Melnick L., Sherman F.;
RT   "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
RT   of Saccharomyces cerevisiae share a common ancestry.";
RL   J. Mol. Biol. 233:372-388(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169868;
RA   Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA   Araujo R., Aviles P.S., Berno A., Brennan T., Carpenter J., Chen E.,
RA   Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA   Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA   Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA   Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA   Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL   Nature 387:78-81(1997).
CC   -!- SIMILARITY: SOME, TO YEAST YGR189C.
```

```
DR   EMBL; L22173; AAA34941.1; -.
DR   EMBL; S65964; AAD13975.1; -.
DR   EMBL; S66130; AAB28444.1; -.
DR   EMBL; U18779; AAB65002.1; ALT_INIT.
DR   PIR; S30839; S30839.
DR   HSSP; P23904; 1AJO.
DR   SGD; S0000766; UTR2.
```

```
DR   GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR   GO; GO:0000144; C:septin ring (sensu Saccharomyces); IDA.
DR   InterPro; IPR000757; Glyco_hydro_16.
DR   Pfam; PF00722; Glyco_hydro_16; 1.
FT   DOMAIN       234    322       SER-RICH.
FT   DOMAIN       269    283       POLY-SER.
FT   CONFLICT      10     10       L -> V (IN REF. 1).
FT   CONFLICT     171    171       A -> R (IN REF. 1).
FT   CONFLICT     234    234       S -> C (IN REF. 1).
SQ   SEQUENCE    347 AA;  36692 MW;  1E8AFB862C4B328 CRC64;

Query Match             3.1%;   Score 11;   DB 1;   Length 347;
Best Local Similarity 100.0%;   Pred. No. 0.0062;
Matches  11;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   325 SSSSSSSSSST 335
             ||||||||||
Db   274 SSSSSSSSSST 284

RESULT 4
WSC2_YEAST
ID   WSC2_YEAST     STANDARD;      PRT;   503 AA.
AC   P53832;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cell wall integrity and stress response component 2 precursor.
GN   WSC2 OR YNL283C OR N0583.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Messenguy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,
RA   Glansdorf N.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 WSC domain.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; Z71559; CAA96195.1; -.
DR   PIR; S63259; S63257.
DR   SGD; S0005227; WSC2.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0004888; F:transmembrane receptor activity; IGI.
DR   GO; GO:0007047; P:cell wall organization and biogenesis; IGI.
DR   GO; GO:0009408; P:response to heat; IGI.
DR   GO; GO:0007266; P:Rho protein signal transduction; IGI.
DR   InterPro; IPR002889; WSC.
DR   Pfam; PF01822; WSC; 1.
DR   SMART; SM00321; WSC; 1.
KW   Cell wall; Transmembrane; Glycoprotein; Signal.
FT   SIGNAL         1     23       POTENTIAL.
FT   CHAIN         24    503       CELL WALL INTEGRITY AND STRESS RESPONSE
FT                                 COMPONENT 2.
FT   TRANSMEM     326    346       POTENTIAL.
FT   DOMAIN       121    319       SER/THR-RICH.
FT   CARBOHYD     394    394       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     429    429       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     475    475       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     483    483       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     498    498       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     499    499       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    503 AA;  52292 MW;  F2392A73C5CBAB50 CRC64;

Query Match             3.1%;   Score 11;   DB 1;   Length 503;
Best Local Similarity 100.0%;   Pred. No. 0.0085;
Matches  11;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   325 SSSSSSSSSST 335
             |||||||||||
Db   183 SSSSSSSSSST 193

RESULT 5
PHR2_CANAL
ID   PHR2_CANAL     STANDARD;      PRT;   546 AA.
AC   O13318;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   pH responsive protein 2 precursor (pH-regulated protein 2).
GN   PHR2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SC5314;
RX   MEDLINE=97459694; PubMed=9315654;
RA   Muehlschlegel F.A., Fonzi W.A.;
RT   "PHR2 of Candida albicans encodes a functional homolog of the pH-
RT   regulated gene PHR1 with an inverted pattern of pH-dependent
RT   expression.";
RL   Mol. Cell. Biol. 17:5960-5967(1997).
CC   -!- FUNCTION: REQUIRED FOR APICAL CELL GROWTH AND PLAYS AN ESSENTIAL
CC       ROLE IN MORPHOGENESIS. MAY BE INTEGRAL TO THE PATHOGENIC ABILITY
CC       OF THE ORGANISM (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- INDUCTION: REPRESSED AT PH VALUES ABOVE 6 AND PROGRESSIVELY
CC       INDUCED AT MORE ACIDIC PH VALUES.
CC   -!- SIMILARITY: Belongs to the GAS1 family.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AF011386; AAB80716.1; -.
DR   InterPro; IPR004886; GAS1.
DR   Pfam; PF03198; GAS1; 1.
KW   Glycoprotein; Membrane; GPI-anchor; Signal.
FT   SIGNAL         1     22       POTENTIAL.
FT   CHAIN         23      ?       PH RESPONSIVE PROTEIN 2.
FT   PROPEP         ?    546       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       442    523       SER-RICH.
FT   DOMAIN       493    515       POLY-SER.
FT   CARBOHYD      40     40       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      59     59       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     149    149       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     410    410       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    546 AA;  58979 MW;  951EEODB9489CED6 CRC64;

Query Match             3.1%;   Score 11;   DB 1;   Length 546;
Best Local Similarity 100.0%;   Pred. No. 0.0091;
Matches  11;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   324 SSSSSSSSSSS 334
             |||||||||||
Db   496 GSSSSSSSSSS 506

RESULT 6
```

AF9_HUMAN
```
ID   AF9_HUMAN      STANDARD;      PRT;    568 AA.
AC   P42568;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   AF-9 protein.
GN   MLLT3 OR AF9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281633; PubMed=8506309;
RA   Nakamura T., Alder H., Gu Y., Prasad R., Canaani O., Kamada N.,
RA   Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M., Canaani E.;
RT   "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in
RT   acute leukemia share sequence homology and/or common motifs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- DISEASE: Involved in acute leukemias by a chromosomal
CC       translocation t(9;11)(p22;q23) that involves MLLT3 and MLL/HRX.
CC       The result is a rogue activator protein.
CC   -!- SIMILARITY: TO HUMAN MLLT1/ENL.
CC   -!- SIMILARITY: TO YEAST TRANSCRIPTION INITIATION FACTOR TFIIF SMALL
CC       SUBUNIT (TFG3/ANC1).
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/AF9.html".
```
```
DR   EMBL; L13744; AAA58361.1; -.
DR   PIR; I39411; I39411.
DR   Genew; HGNC:7136; MLLT3.
DR   MIM; 159558; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0007048; P:oncogenesis; TAS.
DR   InterPro; IPR005033; YEATS.
DR   Pfam; PF03366; YEATS; 1.
KW   Transcription regulation; Activator; Nuclear protein;
KW   Chromosomal translocation; Proto-oncogene.
FT   DOMAIN       149    194       POLY-SER.
FT   DOMAIN       295    300       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN       383    391       POLY-SER.
FT   DOMAIN       466    469       POLY-PRO.
FT   SITE         375    375       MLL FUSION POINT (IN ACUTE MYELOID
FT                                 LEUKEMIA PATIENT CO).
FT   SITE         481    481       MLL FUSION POINT (IN ACUTE MYELOID
FT                                 LEUKEMIA PATIENT F1).
SQ   SEQUENCE   568 AA;  63367 MW;  8802DE5E88A413C2 CRC64;

Query Match             3.1%;  Score 11;  DB 1;  Length 568;
Best Local Similarity   100.0%;  Pred. No. 0.0095;
Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      325 SSSSSSSSST 335
            |||||||||||
Db      181 SSSSSSSSST 191
```

RESULT 7
RN12_HUMAN
```
ID   RN12_HUMAN     STANDARD;      PRT;    624 AA.
AC   Q9NVW2; Q9Y598;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RING finger protein 12 (LIM domain interacting RING finger protein)
DE   (RING finger LIM domain-binding protein) (R-LIM) (NY-REN-43 antigen).
GN   RNF12 OR RLIM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20469411; PubMed=11013082;
RA   Ostendorff H.P., Bossenz M., Mincheva A., Copeland N.G., Gilbert D.J.,
RA   Jenkins N.A., Lichter P., Bach I.;
RT   "Functional characterization of the gene encoding RLIM, the
RT   corepressor of LIM homeodomain transcription factors.";
RL   Genomics 69:120-130(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Renal cell carcinoma;
RX   MEDLINE=99438124; PubMed=10508479;
RA   Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA   Jongeneel C.V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA   Old L.J.;
RT   "Antigens recognized by autologous antibody in patients with renal-
RT   cell carcinoma.";
RL   Int. J. Cancer 83:456-464(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ACTS AS A NEGATIVE CO-REGULATOR FOR LIM HOMEODOMAIN
CC       TRANSCRIPTION FACTORS. VIA THE RECRUITMENT OF THE SIN3A/HISTONE
CC       DEACETYLASE COREPRESSOR COMPLEX.
CC   -!- SUBUNIT: ASSOCIATES WITH LIM/HOMEOBOX FACTORS.
CC   -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC       in position 134 and 142.
```
```
DR   EMBL; AJ271670; CAC14228.1; -.
DR   EMBL; AF155109; AAD42875.1; ALT_FRAME.
DR   EMBL; AK001334; BAA91632.1; -.
DR   Genew; HGNC:13429; RNF12.
DR   MIM; 300379; -.
DR   GO; GO:0017053; C:transcriptional repressor complex; NAS.
DR   GO; GO:0003714; F:transcriptional co-repressor activity; NAS.
DR   GO; GO:0016481; P:negative regulation of transcription, NAS.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Transcription regulation; Zinc-finger.
FT   DOMAIN       422    506       SER-RICH.
FT   ZN_FING      570    611       RING-TYPE.
FT   DOMAIN       453    481       POLY-SER.
FT   DOMAIN       500    506       POLY-SER.
FT   CONFLICT     126    126       S -> C (IN REF. 1).
```

```
FT  CONFLICT   134   134    N -> D (IN REF. 3).
FT  CONFLICT   144   145    YS -> NR (IN REF. 3).
FT  CONFLICT   418   418    Y -> H (IN REF. 3).
SQ  SEQUENCE   624 AA;  68527 MW;  DE3ADE09ACACBCF8 CRC64;

Query Match                3.1%;   Score 11;  DB 1;  Length 624;
Best Local Similarity     100.0%;  Pred. No. 0.01;
Matches    11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSS 334
          ||||||||||
Db    456 GSSSSSSSSS 466

RESULT 8
HWP1_CANAL
ID   HWP1_CANAL   STANDARD;   PRT;   634 AA.
AC   P46593; O13424; P87019;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hyphal wall protein 1 (Cell elongation protein 2).
GN   HWP1 OR ECE2
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SC5314;
RX   MEDLINE=96198091; PubMed=8626424;
RA   Staab J.F., Ferrer C.A., Sundstrom P.;
RT   "Developmental expression of a tandemly repeated, proline-and
RT   glutamine-rich amino acid motif on hyphal surfaces on Candida
RT   albicans.";
RL   J. Biol. Chem. 271:6298-6305(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SC5314;
RA   Sharkey L.L., Saporito-Irwin S.M., Fonzi W.A.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-225 FROM N.A.
RC   STRAIN=SC5314;
RX   MEDLINE=96198091; PubMed=8626424;
RA   Staab J.F., Ferrer C.A., Sundstrom P.R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HYPHAL SURFACE ANTIGEN.
CC   -!- SUBCELLULAR LOCATION: HYPHAL SURFACE.
CC   -!- TISSUE SPECIFICITY: FOUND IN HYPHAL BUT NOT YEAST FORMS.
CC   -!- PTM: MAY BE O-GLYCOSYLATED.
CC   --------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------
DR   EMBL; U64206; AAC96368.1; -.
DR   EMBL; AF001978; AAB64014.1; -.
DR   EMBL; U29369; AAC49209.1; -.
KW   Repeat; Antigen.
FT   DOMAIN     46   187   14 X 10 AA TANDEM REPEATS OF [EVIQ]-P-
FT                         [CDT]-D-[YNW]-P-[PQ]-[QI]-[QP]-[QDN].
FT   REPEAT     46    58    1 (APPROXIMATE).
FT   REPEAT     59    69    2 (APPROXIMATE).
FT   REPEAT     70    81    3 (APPROXIMATE).
FT   REPEAT     82    91    4.
FT   REPEAT     92   101    5.
FT   REPEAT    102   111    6.
FT   REPEAT    112   121    7.
FT   REPEAT    122   131    8.
FT   REPEAT    132   141    9.
FT   REPEAT    142   151   10.
FT   REPEAT    152   161   11.
FT   REPEAT    162   171   12.
FT   REPEAT    172   179   13 (INCOMPLETE).
FT   REPEAT    180   187   14 (INCOMPLETE).
FT   DOMAIN     53    58    POLY-GLN.
FT   DOMAIN     65    69    POLY-GLN.
FT   DOMAIN    204   207    POLY-THR.
FT   DOMAIN    208   220    POLY-SER.
FT   DOMAIN    291   296    POLY-THR.
FT   DOMAIN    304   307    POLY-THR.
FT   DOMAIN    398   401    POLY-VAL.
FT   CONFLICT  141   141    D -> DVPCDNPFQPD (IN REF. 3).
FT   CONFLICT  296   296    T -> A (IN REF. 2).
FT   CONFLICT  441   441    S -> P (IN REF. 2).
FT   CONFLICT  492   492    P -> S (IN REF. 2).
SQ   SEQUENCE  634 AA;  65372 MW;  F841347576BA6376 CRC64;

Query Match                3.1%;   Score 11;  DB 1;  Length 634;
Best Local Similarity     100.0%;  Pred. No. 0.01;
Matches    11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    325 SSSSSSSSSST 335
          |||||||||||
Db    211 SSSSSSSSSST 221

RESULT 9
FIB2_PETMA
ID   FIB2_PETMA   STANDARD;   PRT;   641 AA.
AC   P33573;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Fibrinogen alpha-2 chain precursor [Contains: Fibrinopeptide A].
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92196058; PubMed=1549566;
RA   Pan Y., Doolittle R.F.;
RT   "cDNA sequence of a second fibrinogen alpha chain in lamprey: an
RT   archetypal version alignable with full-length beta and gamma
RT   chains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:2066-2070(1992).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT. THE SOFT CLOT IS
CC       CONVERTED INTO THE HARD CLOT BY FACTOR XIIIA WHICH CATALYZES THE
CC       EPSILON-(GAMMA-GLUTAMYL)LYSINE CROSS-LINKING BETWEEN GAMMA CHAINS
CC       (STRONGER) AND BETWEEN ALPHA CHAINS (WEAKER) OF DIFFERENT
CC       MONOMERS.
CC   -!- SIMILARITY: Contains 1 fibrinogen C-terminal domain.
CC   --------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------
DR   EMBL; M84565; AAA73183.1; -.
```

```
DR   EMBL; M84482; AAA49264.1; -.
DR   HSSP; P02671; 1FZD.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   Pfam; PF00147; fibrinogen_C; 1.
DR   SMART; SM00186; FBG; 1.
DR   PROSITE; PS00186; FIBRIN AG C DOMAIN; 1.
KW   Blood coagulation; Plasma; Signal; Repeat.
FT   SIGNAL        1     23       POTENTIAL.
FT   PEPTIDE      24     31       FIBRINOPEPTIDE A (BY SIMILARITY).
FT   CHAIN        32    641       FIBRINOGEN ALPHA-2 CHAIN.
FT   SITE         31     32       CLEAVAGE (BY THROMBIN; RELEASE
FT                                FIBRINOPEPTIDE A).
FT   DOMAIN      293    316       SER-RICH.
FT   DISULFID     45     45       INTERCHAIN (WITH ALPHA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID     54     54       INTERCHAIN (WITH BETA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID     63     63       INTERCHAIN (WITH GAMMA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID     67     67       INTERCHAIN (WITH BETA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID    179    179       INTERCHAIN (WITH GAMMA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID    183    183       INTERCHAIN (WITH BETA CHAIN)
FT                                (BY SIMILARITY).
FT   DISULFID    404    435       BY SIMILARITY.
FT   DISULFID    571    584       BY SIMILARITY.
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    397    397       N-LINKED (GLCNAC. . .)  (POTENTIAL).
FT   CARBOHYD    458    458       N-LINKED (GLCNAC. . .)  (POTENTIAL).
SQ   SEQUENCE    641 AA;  70756 MW;  CA991A8DD698BFB5 CRC64;

Query Match             3.1%;
Best Local Similarity   100.0%;
Matches    11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;  Pred. No. 0.01;  Length 641;

Qy      325 SSSSSSSSST 335
            |||||||||||
Db      300 SSSSSSSSST 310

RESULT 10
NP14_RAT
ID   NP14_RAT     STANDARD;      PRT;   704 AA.
AC   P41777;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE   nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE   phosphoprotein 1).
GN   NOLC1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.
RC   TISSUE=liver;
RX   MEDLINE=92323542; PubMed=1623516;
RA   Meier U.T., Blobel G.;
RT   "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL   Cell 70:127-138(1992).
RN   [2]
RP   INTERACTION WITH NOP5 AND FIBRILLARIN.
RX   MEDLINE=20143579; PubMed=10679015;
RA   Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
RT   "Conserved composition of mammalian box H/ACA and box C/D small
RT   nucleolar ribonucleoprotein particles and their interaction with the
RT   common factor Nopp140.";
RL   Mol. Biol. Cell 11:567-577(2000).
CC   -!- FUNCTION: RELATED TO NUCLEOLOGENESIS, MAY PLAY A ROLE IN THE
CC       MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC       AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC       GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC       TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC   -!- SUBUNIT: Interacts with Dkc1/Nap57, Nop5/Nap65 and fibrillarin.
CC   -!- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC       NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC       FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOPLASM TO
CC       A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC   -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC       DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPP140 IS ONE OF THE
CC       MOST PHOSPHORYLATED PROTEINS IN THE CELL.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M94287; AAA41718.1; -.
DR   EMBL; M94288; AAA41719.1; -.
DR   PIR; B42680; B42680.
DR   InterPro; IPR006594; LisH.
DR   InterPro; IPR003993; treacle.
DR   Pfam; PF05022; SRP40_C; 1.
DR   PRINTS; PR01503; TREACLE.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
FT   DOMAIN       10     42       LISH.
FT   DOMAIN       84    570       11 X 12 AA APPROXIMATE REPEATS OF AN
FT                                ACIDIC SERINE CLUSTER.
FT   REPEAT       84     95       ACIDIC SERINE CLUSTER 1.
FT   REPEAT      127    138       ACIDIC SERINE CLUSTER 2.
FT   REPEAT      170    181       ACIDIC SERINE CLUSTER 3.
FT   REPEAT      231    242       ACIDIC SERINE CLUSTER 4.
FT   REPEAT      274    285       ACIDIC SERINE CLUSTER 5.
FT   REPEAT      335    346       ACIDIC SERINE CLUSTER 6.
FT   REPEAT      373    384       ACIDIC SERINE CLUSTER 7.
FT   REPEAT      434    445       ACIDIC SERINE CLUSTER 8.
FT   REPEAT      479    490       ACIDIC SERINE CLUSTER 9.
FT   REPEAT      524    535       ACIDIC SERINE CLUSTER 10.
FT   REPEAT      559    570       ACIDIC SERINE CLUSTER 11.
FT   MOD_RES     567    567       PHOSPHORYLATION (BY CK2).
FT   VARIANT     150    150       MISSING (IN NOPP140B).
SQ   SEQUENCE    704 AA;  73562 MW;  14DF1BF2DB483EA3 CRC64;

Query Match             3.1%;
Best Local Similarity   100.0%;  Pred. No. 0.011;  Length 704;
Matches    11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      324 GSSSSSSSSSS 334
            |||||||||||
Db      225 GSSSSSSSSSS 235

RESULT 11
CT67_MOUSE
ID   CT67_MOUSE   STANDARD;      PRT;   706 AA.
AC   P59114;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   protein C20orf67 homolog.
GN   C20ORF67.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 1 WW domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; BC031431; AAH31431.1; -.
DR   MGD; MGI:2443858; F730014I05Rik.
DR   InterPro; IPR001202; WW_Rsp5_WWP.
DR   Pfam; PF00397; WW; 1.
DR   SMART; SM00456; WW; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; FALSE_NEG.
DR   PROSITE; PS50020; WW_DOMAIN_2; 1.
FT   DOMAIN   43   77   WW.
FT   DOMAIN   675   686   POLY-SER.
SQ   SEQUENCE   706 AA; 80504 MW; 48646523758CF4FD CRC64;

Query Match   3.1%; Score 11; DB 1; Length 706;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
              |||||||||||
Db   674 GSSSSSSSSS 684

RESULT 12
GSQD_ERWCH
ID   GSQD_ERWCH   STANDARD;   PRT;   710 AA.
AC   Q01565;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   General secretion pathway protein D precursor (Pectic enzymes
DE   secretion protein OutD).
GN   OUTD.
OS   Erwinia chrysanthemi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Pectobacterium.
OX   NCBI_TaxID=556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3937;
RX   MEDLINE=93086427; PubMed=1453958;
RA   Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;
RT   "Some of the out genes involved in the secretion of pectate lyases in
RT   Erwinia chrysanthemi are regulated by kdgR.";

RL   Mol. Microbiol. 6:3199-3211(1992).
CC   -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC   EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE
CC   MULTIPLE PECTIC ENZYMES.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Probable).
CC   -!- SIMILARITY: BELONGS TO THE EXED/OUTD/PULD/XPSD FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X65265; CAA46370.1; -.
DR   PIR; S28014; S28014.
DR   InterPro; IPR001775; Bac_GSPD.
DR   InterPro; IPR004846; GSPII/IIIprotein.
DR   InterPro; IPR004845; GSPIIproteinC.
DR   InterPro; IPR005644; NolW-like.
DR   Pfam; PF00263; GSPII_III; 1.
DR   Pfam; PF03958; GSPII_III_N; 3.
DR   PRINTS; PR00811; BCTERIALGSPD.
DR   PROSITE; PS00875; T2SP_D; 1.
KW   Transport; Outer membrane; Signal.
FT   SIGNAL   1   27   POTENTIAL.
FT   CHAIN   28   710   GENERAL SECRETION PATHWAY PROTEIN D.
FT   DOMAIN   288   353   GLY/SER-RICH.
SQ   SEQUENCE   710 AA; 76213 MW; 156E84CC50CD54FA CRC64;

Query Match   3.1%; Score 11; DB 1; Length 710;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
              |||||||||||
Db   328 GSSSSSSSSS 338

RESULT 13
NFL1_MOUSE
ID   NFL1_MOUSE   STANDARD;   PRT;   741 AA.
AC   Q61985; O70234;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nuclear factor erythroid 2 related factor 1 (NF-E2 related factor 1)
DE   (NFE2-related factor 1) (Nuclear factor, erythroid derived 2, like 1).
GN   NFE2L1 OR NRF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95278942; PubMed=7759107;
RA   McKie J., Johnstone K., Mattei M.-G., Scambler P.;
RT   "Cloning and mapping of murine Nfe2l1.";
RL   Genomics 25:716-719(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX   MEDLINE=98248571; PubMed=9580677;
RA   Prieschl E.E., Novotny V., Csonga R., Jaksche D., Elbe-Burger A.,
RA   Thumb W., Auer M., Stingl G., Baumruker T.;
RT   "A novel splice variant of the transcription factor Nrf1 interacts
RT   with the TNFalpha promoter and stimulates transcription.";
RL   Nucleic Acids Res. 26:2291-2297(1998).
CC   -!- FUNCTION: THE SHORT ISOFORM INTERACTS WITH THE EXTENDED KAPPA 3
CC   SITE OF THE TNF ALPHA PROMOTER AFTER FC GAMMA RIII STIMULATION AND
CC   PARTICIPATES IN THE INDUCTION OF THIS CYTOKINE. THE LONG ISOFORM
CC   IS EITHER INACTIVE OR REPRESSES THE TRANSCRIPTIONAL ACTIVATION.

```
CC   -!- SUBUNIT: HETERODIMER (PROBABLE). IT MAY FORM A HETERODIMER WITH
CC       SEVERAL TRANSCRIPTION FACTORS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q61985-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q61985-2; Sequence=VSP_000580, VSP_000581;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Widely expressed.
CC   -!- PTM: COULD BE DEPENDENT ON CKII PHOSPHORYLATION FOR BINDING.
CC   -!- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X78709; CAA55362.1; -.
DR   EMBL; AF015881; AAC40108.1; -.
DR   PIR; I48694; I48694.
DR   HSSP; P34707; 1SKN.
DR   MGD; MGI:99421; Nfe2l1.
DR   InterPro; IPR004827; TF_bZIP.
DR   SMART; SM00338; BRlZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
DR   PROSITE; PS00217; BZIP; 1.
KW   DNA-binding; Nuclear protein; Phosphorylation; Alternative splicing.
FT   DOMAIN      125    259       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      414    447       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      467    486       POLY-SER.
FT   DNA_BIND    628    643       BASIC MOTIF (BY SIMILARITY).
FT   DOMAIN      651    673       LEUCINE-ZIPPER (BY SIMILARITY).
FT   DOMAIN      730    737       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARSPLIC      1    291       Missing (in isoform Short).
FT                                /FTId=VSP_000580.
FT   VARSPLIC    447    583       Missing (in isoform Short).
FT                                /FTId=VSP_000581.
FT   CONFLICT    318    318       S -> T (IN REF. 2).
FT   CONFLICT    387    387       P -> L (IN REF. 2).
SQ   SEQUENCE   741 AA;  81545 MW;  C01B89DD26E7CDEF CRC64;

Query Match             3.1%;  Score 11;  DB 11;  Length 741;
Best Local Similarity  100.0%;  Pred. No. 0.012;
Matches   11  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    324 GSSSSSSSSS 334
          ||||||||||
Db    466 GSSSSSSSSS 476

RESULT 14
ID   NFL1_HUMAN    STANDARD;      PRT;    772 AA.
AC   Q14494; Q12877; Q96FN6;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Nuclear factor erythroid 2 related factor 1 (NF-E2 related factor 1)
DE   (NFE2-related factor 1) (Nuclear factor, erythroid derived 2, like 1)
DE   (Transcription factor 11) (Transcription factor HBZ17) (Transcription
DE   factor LCR-F1) (Locus control region-factor 1).
GN   NFE2L1 OR NRF1 OR TCF11 OR HBZ17.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=95095252; PubMed=8001966;
RA   Luna L., Johnsen O., Skartlien A.H., Pedeutour F., Turc-Carel C.,
RA   Prydz H., Kolstoe A.-B.;
RT   "Molecular cloning of a putative novel human bZIP transcription
RT   factor on chromosome 17q22.";
RL   Genomics 22:553-562(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 326-772 FROM N.A.
RX   MEDLINE=94310069; PubMed=8036168;
RA   Caterina J.J., Donze D., Sun C.W., Ciavatta D.J., Townes T.M.;
RT   "Cloning and functional characterization of LCR-F1: a bZIP
RT   transcription factor that activates erythroid-specific, human globin
RT   gene expression.";
RL   Nucleic Acids Res. 22:2383-2391(1994).
CC   -!- FUNCTION: ACTIVATES ERYTHROID-SPECIFIC, GLOBIN GENE EXPRESSION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q14494-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q14494-2; Sequence=VSP_000579;
CC   -!- SIMILARITY: Belongs to the bZIP family. CNC subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X77366; CAA54555.1; -.
DR   EMBL; BC010623; AAH10623.1; -.
DR   EMBL; U08853; AAA20466.1; -.
DR   PIR; A49672; A49672.
DR   PIR; A55004; A55004.
DR   HSSP; P34707; 1SKN.
DR   Genew; HGNC:7781; NFE2L1.
DR   MIM; 163260; -.
DR   GO; GO:0003712; F:transcription cofactor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0006783; P:heme biosynthesis; TAS.
DR   GO; GO:0006954; P:inflammatory response; TAS.
DR   GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR   InterPro; IPR004827; TF_bZIP.
DR   SMART; SM00338; BRlZ; 1.
DR   PROSITE; PS00217; BZIP; 1.
```

```
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Alternative splicing.
FT   DOMAIN       125    288       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       496    517       POLY-SER.
FT   DNA_BIND     659    674       BASIC MOTIF.
FT   DOMAIN       682    704       LEUCINE-ZIPPER.
FT   VARSPLIC     242    271       Missing (in isoform 2).
FT                                 /FTId=VSP_000579.
SQ   SEQUENCE    772 AA;  84703 MW;  C868807C6046BEF5 CRC64;

Query Match    3.1%;  Score 11;  DB 1;  Length 772;
Best Local Similarity   100.0%;  Pred. No. 0.012;
Matches  11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSS 334
          |||||||||||
Db    495 GSSSSSSSSS 505


RESULT 15
ID   GUNA_PSEFL     STANDARD;      PRT;     962 AA.
AC   P10476;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE   (Cellulase) (EGA).
GN   CELA.
OS   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sp. Cellulosa;
RX   MEDLINE=89127129; PubMed=2851699;
RA   Hall J., Gilbert H.J.;
RT   "The nucleotide sequence of a carboxymethylcellulase gene from
RT   Pseudomonas fluorescens subsp. cellulosa.";
RL   Mol. Gen. Genet. 213:112-117(1988).
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC       linkages in cellulose, lichenin and cereal beta-D-glucans.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY E (FAMILY 9 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC       domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X12570; CAA31082.1; -.
DR   HSSP; P07986; 1EXG.
DR   InterPro; IPR001919; Bac_celose-bind.
DR   InterPro; IPR002883; CBD_5.
DR   InterPro; IPR001701; Glyco_hydro_9.
DR   InterPro; IPR004197; Glyco_hydro_9Ig.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF02013; CBM_10; 1.
DR   Pfam; PF00553; CBM_2; 1.
DR   Pfam; PF02927; celD_N; 1.
DR   Pfam; PF00759; Glyco_hydro_9; 1.
DR   Pfam; PF08801; PKD; 1.
DR   SMART; SM00637; CBD_II; 1.
DR   SMART; SM00089; PKD; 1.
DR   PROSITE; PS00561; CBD_BACTERIAL; 1.
DR   PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR   PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL         1     32       POTENTIAL.
FT   CHAIN         33    962       ENDOGLUCANASE A.
FT   ACT_SITE     523    523       BY SIMILARITY.
FT   ACT_SITE     573    573       BY SIMILARITY.
FT   ACT_SITE     582    582       BY SIMILARITY.
FT   DOMAIN       608    664       SER-RICH.
FT   DOMAIN       823    859       SER-RICH (LINKER).
FT   DOMAIN       866    962       CELLULOSE-BINDING (BY SIMILARITY).
FT   DISULFID     866    961       BY SIMILARITY.
SQ   SEQUENCE    962 AA;  100290 MW;  9E93FE1EAC21A898 CRC64;

Query Match    3.1%;  Score 11;  DB 1;  Length 962;
Best Local Similarity   100.0%;  Pred. No. 0.015;
Matches  11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSS 334
          |||||||||||
Db    607 GSSSSSSSSS 617


Search completed: February 19, 2004, 19:37:49
Job time : 19 secs
```

```
                GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February 19, 2004, 19:35:01 ; Search time 40 Seconds
                                          (without alignments)
                            2322.473 Million cell updates/sec

Title:          US-10-041-615-34
Perfect score:  360
Sequence:       1 MIKLGPYFTPPTKIKTKIKD.........SETPSITQARGSMFLAEHVV 360

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       830525 seqs, 258052604 residues

Word size :     10

Total number of hits satisfying chosen parameters:     894

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      SPTREMBL 23:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*
                15: sp_rvirus:*
                16: sp_bacteriap:*
                17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	3.6	311	5	Q9V684	Q9v684 drosophila
2	13	3.3	250	5	Q8MN67	Q8mn67 dictyosteli
3	12	3.3	580	5	Q9GUB3	Q9gub3 dictyosteli
4	12	3.3	624	5	Q8NIV1	Q8niv1 neurospora
5	12	3.3	967	5	Q95U13	Q95u13 drosophila
6	12	3.3	1280	5	Q9VWC2	Q9vwc2 drosophila
7	12	3.3	2529	5	Q9VWC0	Q9vwc0 drosophila
8	11	3.1	50	5	Q9NDB3	Q9ndb3 cryptospori
9	11	3.1	59	5	Q9NCR6	Q9ncr6 cryptospori
10	11	3.1	68	5	Q9NCS0	Q9ncs0 cryptospori
11	11	3.1	104	5	Q9NCR8	Q9ncr8 cryptospori
12	11	3.1	107	5	Q9NCR9	Q9ncr9 cryptospori
13	11	3.1	109	5	Q9U2E9	Q9u2e9 caenorhabdi
14	11	3.1	131	5	Q9NCS2	Q9ncs2 cryptospori
15	11	3.1	134	16	Q8NQN3	Q8nqn3 corynebacte
16	11	3.1	136	12	O73451	O73451 human papil
17	11	3.1	138	5	Q8N0K1	Q8n0k1 cryptospori
18	11	3.1	141	5	Q8N0J1	Q8n0j1 cryptospori
19	11	3.1	145	11	Q8BY58	Q8by58 mus musculu
20	11	3.1	146	10	Q8GYN0	Q8gyn0 arabidopsis
21	11	3.1	149	11	Q8C5W6	Q8c5w6 mus musculu
22	11	3.1	153	5	Q8N0I2	Q8n0i2 cryptospori
23	11	3.1	156	5	Q8N0I1	Q8n0i1 cryptospori
24	11	3.1	157	5	Q8N0H0	Q8n0h0 cryptospori
25	11	3.1	158	5	Q8N0H1	Q8n0h1 cryptospori
26	11	3.1	158	5	Q8N0G5	Q8n0g5 cryptospori
27	11	3.1	158	5	Q8N0H2	Q8n0h2 cryptospori
28	11	3.1	158	5	Q8N0I0	Q8n0i0 cryptospori
29	11	3.1	159	5	Q8N0G9	Q8n0g9 cryptospori
30	11	3.1	159	5	Q8N0I3	Q8n0i3 cryptospori
31	11	3.1	162	10	Q9M0B4	Q9m0b4 arabidopsis
32	11	3.1	165	5	Q8N0G8	Q8n0g8 cryptospori
33	11	3.1	170	5	Q8N0H9	Q8n0h9 cryptospori
34	11	3.1	170	5	Q8N0H4	Q8n0h4 cryptospori
35	11	3.1	173	5	Q8N0I6	Q8n0i6 cryptospori
36	11	3.1	173	5	Q8N0J0	Q8n0j0 cryptospori
37	11	3.1	175	5	Q8N0I7	Q8n0i7 cryptospori
38	11	3.1	178	5	Q8N0K2	Q8n0k2 cryptospori
39	11	3.1	178	5	Q8N0I9	Q8n0i9 cryptospori
40	11	3.1	178	10	Q9LRJ8	Q9lrj8 arabidopsis
41	11	3.1	178	10	Q8LBR9	Q8lbr9 arabidopsis
42	11	3.1	179	5	Q8N0I4	Q8n0i4 cryptospori
43	11	3.1	181	5	Q8N0I5	Q8n0i5 cryptospori
44	11	3.1	193	10	Q43454	Q43454 glycine max
45	11	3.1	193	11	Q9D6J0	Q9d6j0 mus musculu

ALIGNMENTS

```
RESULT 1
Q9V684
ID   Q9V684          PRELIMINARY;   PRT;   311 AA.
AC   Q9V684;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG13169 protein.
GN   CG13169.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Murphy D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard V., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003823; AAF58547.1; -.
DR   FlyBase; FBgn0033704; CG13169.
SQ   SEQUENCE   311 AA; 33572 MW; B75163C3AF5BACE3 CRC64;

Query Match   3.6%; Score 13; DB 5; Length 311;
  Best Local Similarity   100.0%; Pred. No. 0.00015;
  Matches   13; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSSSTP 336
              |||||||||||||
Db   188 GSSSSSSSSSSTP 200

RESULT 2
Q8MN67
ID   Q8MN67;   PRELIMINARY;   PRT;   250 AA.
AC   Q8MN67;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117079; AAM43726.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   250 AA; 27881 MW; 163B7C7B3A50264B CRC64;

Query Match   3.3%; Score 12; DB 5; Length 250;
  Best Local Similarity   100.0%; Pred. No. 0.0013;
  Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSTP 336
              |||||||||||
Db   122 SSSSSSSSSTP 133

RESULT 3
Q9GUB3
ID   Q9GUB3;   PRELIMINARY;   PRT;   580 AA.
AC   Q9GUB3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Myb protein.
OS   Dictyostelium discoideum (slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;

RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99265952; PubMed=10331990;
RA   Guo K., Anjard C., Harwood A., Kim H.J., Newell P.C., Gross J.D.;
RT   "A myb-related protein required for culmination in Dictyostelium.";
RL   Development 126:2811-2822(1999).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.
DR   EMBL; AF098507; AAG15274.1; -.
DR   HSSP; P06876; 1MBG.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF00249; myb_DNA-binding; 3.
DR   SMART; SM00717; SANT; 3.
DR   PROSITE; PS00037; MYB_1; 1.
DR   PROSITE; PS50090; MYB_3; 3.
KW   DNA-binding; Nuclear protein.
SQ   SEQUENCE   580 AA; 66019 MW; F02D9FACB79F3C43 CRC64;

Query Match   3.3%; Score 12; DB 5; Length 580;
  Best Local Similarity   100.0%; Pred. No. 0.0027;   Indels   0; Gaps   0;
  Matches   12; Conservative   0; Mismatches   0;

Qy   324 GSSSSSSSSSST 335
              ||||||||||||
Db   191 GSSSSSSSSSST 202

RESULT 4
Q8NIV1
ID   Q8NIV1;   PRELIMINARY;   PRT;   624 AA.
AC   Q8NIV1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   B13H18.080.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL807374; CAD37056.1; -.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF01391; Collagen; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   624 AA; 58999 MW; 73FED5782242E4D9 CRC64;

Query Match   3.3%; Score 12; DB 3; Length 624;
  Best Local Similarity   100.0%; Pred. No. 0.0029;   Indels   0; Gaps   0;
  Matches   12; Conservative   0; Mismatches   0;

Qy   325 SSSSSSSSSSTP 336
              ||||||||||||
Db   146 SSSSSSSSSSTP 157

RESULT 5
Q95U13
ID   Q95U13;   PRELIMINARY;   PRT;   967 AA.
AC   Q95U13;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GH14389p.
GN   CG32529 OR CG11936 OR CG15619.

```
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkeley;
RA  Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA  Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA  Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA  Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA  Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL  Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AY058379; AAL13608.1; -.
DR  FlyBase; FBgn0052529; CG32529.
DR  InterPro; IPR001025; BAH.
DR  Pfam; PF01426; BAH; 1.
DR  SMART; SM00439; BAH; 1.
SQ  SEQUENCE 967 AA; 105814 MW; 7DFD5B4521A3DC2A CRC64;

Query Match      3.3%;    Score 12;  DB 5; Length 967;
Best Local Similarity  100.0%;  Pred. No. 0.0043;
Matches  12;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  324 GSSSSSSSSST 335
        |||||||||||
Db   56 GSSSSSSSSST 67


RESULT 6
Q9VWC2
ID  Q9VWC2          PRELIMINARY;     PRT;  1280 AA.
AC  Q9VWC2;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  CG32529 protein.
GN  CG32529 OR CG11936 OR CG15619.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkeley;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA  Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA  Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA  Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA  Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA  Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA  Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA  Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA  Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA  McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA  Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA  Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA  Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA  Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT  "Sequencing of Drosophila melanogaster genome.";
RL  Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RA  Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA  Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA  Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA  Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA  Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA  Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA  Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT  "Annotation of Drosophila melanogaster genome.";
RL  Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN  [4]
RP  SEQUENCE FROM N.A.
RA  Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN  [5]
RP  SEQUENCE FROM N.A.
RA  FlyBase;
RL  Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AE003513; AAF49024.2; -.
DR  FlyBase; FBgn0052529; CG32529.
DR  InterPro; IPR001025; BAH.
DR  Pfam; PF01426; BAH; 1.
DR  SMART; SM00439; BAH; 1.
SQ  SEQUENCE 1280 AA; 138439 MW; 32331F1FD60E75DB CRC64;

Query Match      3.3%;    Score 12;  DB 5; Length 1280;
Best Local Similarity  100.0%;  Pred. No. 0.0054;
Matches  12;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  324 GSSSSSSSSST 335
        |||||||||||
Db  369 GSSSSSSSSST 380


RESULT 7
Q9VWC0
ID  Q9VWC0          PRELIMINARY;     PRT;  2529 AA.
AC  Q9VWC0;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  CG32529 protein.
```

```
GN   CG32529 OR CG11936 OR CG15619.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20195006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Dietz S.M.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.S., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
```

```
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003513; AAF49026.2; -.
DR   FlyBase; FBgn0052529; CG32529.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001025; BAH.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   SMART; SM00439; BAH; 1.
SQ   SEQUENCE   2529 AA;   270675 MW;   C93B1726EE7547D4 CRC64;

Query Match             3.3%;  Score 12;  DB 5;  Length 2529;
Best Local Similarity   100.0%;  Pred. No. 0.0098;
Matches   12;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSST 335
          |||||||||||
Db   1618 GSSSSSSSSST 1629

RESULT 8
Q9NDB3
ID   Q9NDB3                  PRELIMINARY;        PRT;    50 AA.
AC   Q9NDB3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   15 kDa glycoprotein gp15 (Fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GCH1seqA; TISSUE=Oocyst;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF164508; AAF78362.1; -.
FT   NON_TER      1      1
FT   NON_TER     50     50
SQ   SEQUENCE   50 AA;   4705 MW;   D87D5DFE8811C421 CRC64;

Query Match             3.1%;  Score 11;  DB 5;  Length 50;
Best Local Similarity   100.0%;  Pred. No. 0.0035;
Matches   11;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    324 GSSSSSSSSSS 334
          ||||||||||
Db     24 GSSSSSSSSSS 34

RESULT 9
Q9NCR6
ID   Q9NCR6                  PRELIMINARY;        PRT;    59 AA.
AC   Q9NCR6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Gp15 antigen (Fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
```

OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HGM10;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF178697; AAF81991.1; -.
FT   NON_TER   1   1
FT   NON_TER   59   59
SQ   SEQUENCE   59 AA;   5742 MW;   D9173B91CB482AA3 CRC64;

Query Match   3.1%; Score 11; DB 5; Length 59;
Best Local Similarity   100.0%; Pred. No. 0.004;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
            ||||||||||
Db   29 GSSSSSSSSS 39

RESULT 10
Q9NCS0
ID   Q9NCS0   PRELIMINARY;   PRT;   68 AA.
AC   Q9NCS0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Gp15 antigen (Fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HS1;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF178693; AAF81987.1; -.
FT   NON_TER   1   1
FT   NON_TER   68   68
SQ   SEQUENCE   68 AA;   6551 MW;   430C6lCA4A22799C CRC64;

Query Match   3.1%; Score 11; DB 5; Length 68;
Best Local Similarity   100.0%; Pred. No. 0.0045;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
            ||||||||||
Db   27 GSSSSSSSSS 37

RESULT 11
Q9NCR8
ID   Q9NCR8   PRELIMINARY;   PRT;   104 AA.
AC   Q9NCR8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Gp15 antigen (Fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]

RP   SEQUENCE FROM N.A.
RC   STRAIN=HGM07;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF178695; AAF81989.1; -.
FT   NON_TER   1   1
FT   NON_TER   104   104
SQ   SEQUENCE   104 AA;   9926 MW;   34610E5523B53C2F CRC64;

Query Match   3.1%; Score 11; DB 5; Length 104;
Best Local Similarity   100.0%; Pred. No. 0.0065;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
            ||||||||||
Db   29 GSSSSSSSSS 39

RESULT 12
Q9NCR9
ID   Q9NCR9   PRELIMINARY;   PRT;   107 AA.
AC   Q9NCR9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Gp15 antigen (Fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HS2;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF178694; AAF81988.1; -.
FT   NON_TER   1   1
FT   NON_TER   107   107
SQ   SEQUENCE   107 AA;   10273 MW;   95AACDEF836673A CRC64;

Query Match   3.1%; Score 11; DB 5; Length 107;
Best Local Similarity   100.0%; Pred. No. 0.0067;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
            ||||||||||
Db   29 GSSSSSSSSS 39

RESULT 13
Q9U2E9
ID   Q9U2E9   PRELIMINARY;   PRT;   109 AA.
AC   Q9U2E9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Y46G5A.20 protein.
GN   Y46G5A.20.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wallis J.M.;

RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; AL110485; CAB60362.1; -.
DR   WormPep; Y46G5A.20; CE24293.
SQ   SEQUENCE 109 AA; 11493 MW; 124002FF2443810F CRC64;

Query Match   3.1%; Score 11; DB 5; Length 109;
Best Local Similarity   100.0%; Pred. No. 0.0068;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSST 335
              |||||||||||
Db   40 SSSSSSSSSST 50

RESULT 14
Q9NCS2
ID   Q9NCS2   PRELIMINARY;   PRT;   131 AA.
AC   Q9NCS2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Gp15 antigen (fragment).
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6HMA1-10;
RX   MEDLINE=20316040; PubMed=10858229;
RA   Strong W.B., Gut J., Nelson R.G.;
RT   "Cloning and sequence analysis of a highly polymorphic Cryptosporidium
RT   parvum gene encoding a 60-kilodalton glycoprotein and characterization
RT   of its 15- and 45-kilodalton zoite surface antigen products.";
RL   Infect. Immun. 68:4117-4134(2000).
DR   EMBL; AF178691; AAF81985.1; -.
FT   NON_TER   1   1
FT   NON_TER   131   131
SQ   SEQUENCE 131 AA; 12930 MW; 45BB7D7DC31017BA CRC64;

Query Match   3.1%; Score 11; DB 5; Length 131;
Best Local Similarity   100.0%; Pred. No. 0.008;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   324 GSSSSSSSSS 334
              ||||||||||
Db   29 GSSSSSSSSS 39

RESULT 15
Q8NQN3
ID   Q8NQN3   PRELIMINARY;   PRT;   134 AA.
AC   Q8NQN3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein Cgl1393.
GN   CGL1393.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;

RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005278; BAB98786.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 134 AA; 13002 MW; DA00A9CCF281E271 CRC64;

Query Match   3.1%; Score 11; DB 16; Length 134;
Best Local Similarity   100.0%; Pred. No. 0.0081;
Matches   11; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   325 SSSSSSSSSST 335
              |||||||||||
Db   42 SSSSSSSSSST 52

Search completed: February 19, 2004, 19:38:43
Job time : 43 secs